

© 2006 The Authors
Journal compilation © 2006 Blackwell Publishing Ltd

```
OC Actinobacillus.
OX NCBI_TaxID=715;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISU158;
RX MEDLINE=96110919; PubMed=8557378;
RA Cruz W.T., Nedialkov Y.A., Thacker B.J., Mulks M.H.;
RT "Molecular characterization of a common 48-kilodalton outer membrane
RL protein of Actinobacillus pleuropneumoniae.";
DR Infect. Immun. 64:83-90(1996).
SQ EMBL; U24492; AAC43631.1;
SEQUENCE 449 AA; 48604 MW; 62CBDD17A8435418 CRC64;

alignment_scores:
  Quality: 1583.00      Length: 450
  Ratio: 4.059          Gaps: 3
Percent Similarity: 86.667 Percent Identity: 68.889

alignment_block:
US-09-303-518d-131 x Q44130 ..

Align seg 1/1 to: Q44130 from: 1 to: 449

1 ATGATTAATAATCAAAAAGCTCTAAATCGCCATCGCGGCGACACCGGA 50
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
1 MetIleThrIleLysLysGlyLeuAspLeuProIleAlaGlyThrProAl 17

51 GCAAGTCATTATGACGGCCCGGCCATTACCGAAGTCGGTTCGGCG 100
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
17 aGlnValIleHisAsnGlyAsnThrValAsnGluValAlaMetLeuGlyG 34

101 AAGAATATCGCGATCGCGCCCTCGATCAAAATCAAGAGTGAAGCC 150
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
34 IuGluThrValGlyMetArgProSerMetLysValargLuglyAspVal 50

151 GTCAAAAAGGCGCAAGTCGTGTGTGAAGCAAAAAGATCCGGCGTAGT 200
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67

201 ATTTACTGCGCGCTCAGGCAAAATCGCGCTATTACCGTGGCGAAA 250
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
67 lPheThrAlaProAlaSerGlyThrValValThrIleAsnArgGlyGlu 84

251 AGCGCGTACTTCAGTCAGTCGTGATTCGCGTGAAGGCAACACGAAATC 300
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
84 ysargValLeuGlnSerValValIleLysValLuglyAspGluGlnIle 100

301 GAGTTCGAACGTACGTACCTGAGCGCTGGCGAAAATTGACGACGAAA 350
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
101 ThrPheThrArgTyrGluAlaAlaGlnLeuAlaSerLeuSerAlaGluG 117

351 AGTGGCGCGCAACTGATTAATCAGCTTATGACTCGCTTCGCACCC 400
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
117 nValLysGlnAsnLeuIleGluSerGlyLeuThrPheAlaPheArgThr 134

401 GTCGGTTCAGCAAAATCCCTCGCTAGATCCGAGCGCGTTCGCATCTC 450
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
134 rgProPheSerLysValProAlaLeuAspAlaIleProSerSerIlePhe 150

451 GTCATCGCATGACACCAATCGCTGGCTGGCGACCCCTACGTCATCAT 500
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProGluValValle 167

501 CAAGAAGCGCGCCGACACTTCAACCGCGCTGTGGTATTGAGCGGCC 550
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
167 uLysGluTyrCluThrAspPheLysAspGlyLeuThrValLeuThrargL 184

551 TG.....ACGAAACGTAATCCATGTGTAAAGCAGCAGCGCGACAG 594
||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
184 euPheAsnGlyGlnLysProValTyrLeuCysLysAspAlaAspSerAsn 200

595 GTGCGG...TCTGAAATGCTGCCAATATCGAAACACATGAATTTGCGG 641
```

```
201 IleProLeuSerProAlaIleGluGlyIleThrIleLysSerPheSerGI 217
:::||||| ||| ::::|
642 CCGCATCTCGCGGCTTGAGTGGCAGCAGCATTTTCATTCATCGAGCCAG 691
| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
217 yValHisProAlaGlyLeuValGlyThrHisIleHisPheValAspProv 234
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
692 TCGCGCGCAATAAAGCCGTGTGACCATCAATATATCAAGACCTGATGCT 741
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
234 alGlyAlaThrLysGlnValTrpHisLeuAsnTyrGlnAspValIleAla 250
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
742 ATCGGAGCTTGTGTGTAACAGCGCTCTGCAATACCGAGCGCTGCTGC 791
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
251 IleGlyLysLeuPheThrThrGlyGluLeuPheThrAspArgIleIleSe 267
:::||||| ||| ::::|
792 CTTGGCGCGCTGCAAGTCAACAAACCGCGCTCTTTCGCTACCGTTTGG 841
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
267 rLeuAlaGlyProGlnValLysAsnProArgLeuValArgThrArgLeuG 284
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
842 GTGCGAAGTGTCTCAACTTACCGCGCGCAATTTGGTTGACGCGGACAC 891
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
284 lYAlaAsnLeuSerGlnLeuThrAlaAsnGluLeuAsnAlaGlyGluAsn 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
892 CGCGTGTATTCGGTTCGTTTCAACGCGTGCATTCGACAAAGCGCGCA 941
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301 ArgValIleSerGlySerValLeuSerGlyAlaThrAlaAlaGlyProVa 317
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
942 TGATTATTTGGGACGCTACCAATCAATTCGTTTCCGTTATCGAAGAAGCC 991
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
317 lAspTyrLeuGlyArgTyrAlaLeuGlnValSerValLeuAlaGlyGly 334
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
992 GCACGAAGAGCTGTTGGTGGTTCGGCGCGAGCGGACAAATACCTCC 1041
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
334 rgGluLysGluLeuPheGlyTriPheMetProGlySerAspLysPheSer 350
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1042 ATCAGCGCACCACTCTCGGCAATTTCTTAAACAACTCTTCAAGTT 1091
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
351 lIleThrArgThrValLeuGlyHisPheGlyLys...LysLeuPheAsnPh 366
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1092 CACGACAGCCCTCAACGCGCGCGCGCATCGTACCGTACCGTTCGCTT 1141
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
366 eThrThrAlaValHisGlyGlyGluArgAlaMetValProIleGlyAla 383
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1142 ATGACGCGTAAATCGGTGGACATCTGCTACCTGCTGCTTTTTCGCGCAT 1191
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
383 yrGluArgValMetProLeuAspIleIleProThrLeuLeuLeuArgasp 399
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1192 TTAATCGTCGCGGATACCGACGCGCGCGCTTTGGTTCGTTGGAATT 1241
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
400 LeuAlaAlaGlyAspThrAspSerAlaGlnAsnLeuGlyCysLeuGluLe 416
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1242 GGACGAAGAAGCTCGCTTTGTGACGCTTGTGCTGCGCGGCAATACG 1291
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
416 uAspGluGluAspLeuAlaLeuCysThrTyrValCysProGlyLysAsnA 433
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1292 ATACGCGCGCGCTGTCGCAAGTGTGGAACCAATTCAGAAAGGAGGC 1341
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
433 snTyrGlyProMetLeuArgAlaAlaLeuGluLysIleGluLysGluGly 449

seq_name: sp_bacteria:Q9X6S0
seq_documentation_block:
ID Q9X6S0 PRELIMINARY; PRT; 451 AA.
AC Q9X6S0;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 50 KDA ANTIGEN PGI.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; CFB group; Bacteroidetes; Bacteroidales; Porphyromonadaceae;
OX Porphyromonas
NCBI_TaxID=837;
RN [1]
```

RP SEQUENCE FROM N.A.

RC STRAIN=W50;
 RA Ross B.C., Barr I., Patterson M., Agius C., Rothe L., Margetts M.,
 RT "Porphyromonas gingivalis" polypeptides and nucleic acids."
 RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF144076; AAD33930.1; -.
 SQ SEQUENCE 451 AA: 49825 MW: 3E4CC9E66FESAD74 CRC64;

alignment_scores:

Quality: 648.00 Length: 452
 Ratio: 2.189 Gaps: 7
 Percent Similarity: 65.487 Percent Identity: 34.071

alignment_block:

US-09-303-518D-131 x Q9X6S0 ..

Align seg 1/1 to: Q9X6S0 from: 1 to: 451

```

1  ATGATTAAATCAAAAAGGTCTAAATCTGCCCATCGCGGCGAGACCGGA 50
   :::::||||| :::::||||| :::::|||||
4  ValIleLysThrLysLysGlyLeuAlaLeuAsnLeuLysGlyLysProLe 20
   :::::||||| :::::||||| :::::|||||
51 GCAAGTCATTATGACGGCCCGCCCATTCACCGAAGTC...GGCTTCCTTG 97
   :::::||||| :::::||||| :::::|||||
20  uProGluMetLeuAlaGluProAlaGlnSerProThrTyrAlaValValp 37
   :::::||||| :::::||||| :::::|||||
98  GCGAAGAATATGCGCATCGCGCCCTCGATGAAATCAAGNAGTGA 147
   :::::||||| :::::||||| :::::|||||
37  roAspAspPheGluGlyValIleProLysValThrAlaArgProGlyasp 53
   :::::||||| :::::||||| :::::|||||
148 GCGCTCAAAAAGGCGCAAGTGTGTTGAAGACAAAAGAAATCCGGCGCT 197
   :::::||||| :::::||||| :::::|||||
54  LysValArgAlaGlySerAlaLeuMetHisLysLysAlaTyrProGluMe 70
   :::::||||| :::::||||| :::::|||||
198 AGTATTACTGCGCGCTTCAGGCAAAATCGCGCTATTCACCGTGGCG 247
   :::::||||| :::::||||| :::::|||||
70  tLysPheThrSerProValSerGlyGluValIleAlaValAsnArgGlyA 87
   :::::||||| :::::||||| :::::|||||
248 AAAGCGCGTACTTTCAGTCAGTCGTGATGCGCGTTCAGAGCAACGAA 297
   :::::||||| :::::||||| :::::|||||
87  lalysargLysValLeuSerIleGluValLysProAspGlyLeuAsnGlu 103
   :::::||||| :::::||||| :::::|||||
298 ATGAG...TTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTCAGCAG 344
   :::::||||| :::::||||| :::::|||||
104  TyrGluSerPheProValGlyAspProSerAla.....LeuSerAl 117
   :::::||||| :::::||||| :::::|||||
345 CGAAAAGTGCAGCAACCTGATTCATCAGCTATGAGTACGCTGCTTC 394
   :::::||||| :::::||||| :::::|||||
117  aGluGlnIleLysGluLeuLeuLeuSerSerGlyMetTrpGlyPheIleL 134
   :::::||||| :::::||||| :::::|||||
395 GCACCGCTCGCTTTCAGAAAATCCCTGCGGTAGATGCCGAGCGTTCGCC 444
   :::::||||| :::::||||| :::::|||||
134  yslnArgProTyrAspIleValAlaThrProAspIleAlaProArgasp 150
   :::::||||| :::::||||| :::::|||||
445 ATCTGCTCAATGCGATGACCAACATCGCTGGCTGCGCAGCCCTACGGT 494
   :::::||||| :::::||||| :::::|||||
151  ileTyrIleThrAlaAsnPheThrAlaProLeuAlaProAspPheAspPh 167
   :::::||||| :::::||||| :::::|||||
495 CATCATCAAGAAGCGCGCGAGACTTCAACCGCGCTGTTGGTATTCGA 544
   :::::||||| :::::||||| :::::|||||
167  elleValArgGlyGluGluArgAlaLeuGlnThrAlaIleAspAlaLeuA 184
   :::::||||| :::::||||| :::::|||||
545 GCGCGCTGACCGAAGGTAAATTCATGTGTGTAAAGCAGCAGCGCGCAGAC 594
   :::::||||| :::::||||| :::::|||||
184  lalysLeuThrThrGlyLysValTyrValGlyLeuLysProGlySerSer 200
   :::::||||| :::::||||| :::::|||||
595 GTGCGCTGTAATGCTGCCAATATTCGAAACACATGAATTTGGCGGCC 644
   :::::||||| :::::||||| :::::|||||
201  LeuGlyLeuHisAsnAlaGluIleValGluValHis.....Glypr 214
   :::::||||| :::::||||| :::::|||||
645  GCATCTCTGCGCGCTTGAGTGGCAGCACACATTCATTCAGCAGCGCTCG 694

```

```

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
214 ohisProAlaGlyAsnValGlyValLeuIleAsnHisThrLysProIleA 231
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
695 GCGCGAATAAACCCTGTGGAGCCATCAATTATCAAGACGTGATGCTATC 744
   :::::||||| :::::||||| :::::||||| :::::|||||
231 snArgGlyGluThrValTrpThrLeuLysAlaThrAspLeuIleValIle 247
   :::::||||| :::::||||| :::::||||| :::::|||||
745 GGACGCTTCTGCTAACAGCGCTCTGAATACCGAGCGCGTGGTGCCTT 794
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
248 GlyArgPheLeuLeuThrGlyLysAlaAspPheThrArgMetIleAlaMe 264
   :::::||||| :::::||||| :::::||||| :::::|||||
795 GGGCGCGCTGCAAGTCAACAAACCCTGCTGCTGCTACCGTTCCTGGGT 844
   :::::||||| :::::||||| :::::||||| :::::|||||
264 tThrGlySerAspAlaAlaAlaHisGlyTyrValArgIleMetProGlyC 281
   :::::||||| :::::||||| :::::||||| :::::|||||
845 CGAAGGTGCTCAACTTACCGCGCGCGAATGGTT.....GACGCGAC 888
   :::::||||| :::::||||| :::::||||| :::::|||||
281 ysAsnValPheAlaSerPheProGlyArgLeuThrIleLysGluSerHis 297
   :::::||||| :::::||||| :::::||||| :::::|||||
889 AACCGCGTATTTCGGTTCGTTTGAACGTCGCGTTCGATTCGACAGCGC 938
   :::::||||| :::::||||| :::::||||| :::::|||||
298 GluArgValIleAspGlyAsnValLeuThrGlyLysLysLeuCysGluLy 314
   :::::||||| :::::||||| :::::||||| :::::|||||
939 GCATGATTATTGGGACGCTACCAATCAGATTCCTGTTATTCGAAGAAG 988
   :::::||||| :::::||||| :::::||||| :::::|||||
314 sGluProPheLeuSerAlaArgCysAspGlnIleThrValIleProGluG 331
   :::::||||| :::::||||| :::::||||| :::::|||||
989 GCCGAGC....AAGAGCTGTTCGGTGGTTCGCGCGCAGCGGACAAA 1035
   :::::||||| :::::||||| :::::||||| :::::|||||
331 lyAspValAspGluLeuPheGlyTrpAlaAlaProArgLeuAspGln 347
   :::::||||| :::::||||| :::::||||| :::::|||||
1036 TACTCCATCAGCGCACCTCTCGGCTTCCTA...AAAAACAACACT 1082
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
348 TyrSerMetSerArgAlaTyrPheSerTrpLeuGlnGlyLysAsnLysG 364
   :::::||||| :::::||||| :::::||||| :::::|||||
1083 CTTCAAGTTCACGACGCGTCAACGCGCGCGCGCCCATGTCACCGA 1132
   :::::||||| :::::||||| :::::||||| :::::|||||
364 uTyrValLeuAspAlaArgIleLysGlyGlyGluArgAlaMetIleMet 381
   :::::||||| :::::||||| :::::||||| :::::|||||
1133 TCGGCACTTATGCGCGTAAATGCGGTGGACATCCTGCTGCTGCTTCTT 1182
   :::::||||| :::::||||| :::::||||| :::::|||||
381 erAsnGluTyrAspArgValPheProMetAspIleTyrProGluTyrLeu 397
   :::::||||| :::::||||| :::::||||| :::::|||||
1183 TTGCGCGATTAAATCGTGGGATACCGACGCGCGCGCTTGGCTTG 1232
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
398 LeuLysAlaIleIleAlaPheAspIleAspLysMetGluAspLeuGlyI 414
   :::::||||| :::::||||| :::::||||| :::::|||||
1233 CTTGGAATTGACGAAGAAGACCTCGCTTTGTGCAGCTTCGCTGCGCCG 1282
   :::::||||| :::::||||| :::::||||| :::::|||||
414 eTyrGluValAlaProGluAspPheAlaThrCysGluPheValAspThrS 431
   :::::||||| :::::||||| :::::||||| :::::|||||
1283 GCAATACGAATACGCGCGCTGTTCGCAAGTCTGGAAACCATTCAG 1332
   :::::||||| :::::||||| :::::||||| :::::|||||
431 erLysIleGluLeuGlnArgIleValArgGluGlyLeuAspMetLeuTyr 447
   :::::||||| :::::||||| :::::||||| :::::|||||
1333 AAGGAA 1338
   |||||
448 LysGlu 449

```

seq_name: sp_human:Q9UQ36

seq_documentation_block:

ID Q9UQ36 PRELIMINARY; PRT; 1275 AA.

AC Q9UQ36;

DT 01-MAY-2000 (TReMBLrel. 13, Created)

DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)

DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)

DE RNA BINDING PROTEIN (FRAGMENT).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

[illegible]

Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,

RA Tanaka A., Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. VII.

RT The complete sequences of 100 new cDNA clones from brain which can

RL code for large proteins in vitro.";

RL DNA Res. 4:141-150(1997).

DR EMBL; AB023322; BAA20782.2; -.

FT NON_TER 1

SQ SEQUENCE 1783 AA; 190940 MW; 660302F6FD4179AB CRC64;

alignment_scores:

Quality: 141.00 Length: 446

Ratio: 0.701 Gaps: 17

Percent Similarity: 45.067 Percent Identity: 25.336

alignment_block:

US-09-303-518D-131 x 015038 ..

Align seg 1/1 to: 015038 from: 1 to: 1783

```

32 CCATCGCGGCGAGACCGGAGCAAGTCAATTTATGACGGCCGCGCCATTACC 81
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
752 ProProArgHisArgArgSerProSerValSerProGluProAlaG1 768
82 GAAGTCGGCTTCCTTGGCGAGAATATGCGGCATGCGCCCTCGATGAA 131
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
768 uLysSerArgSerArgArgArgArgSerAlaSerSerProArgThrL 785
132 AATCAAGGAAGTGAAGCCGTCAAAAAGGCAAGTGTGTTGAAGACA 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
785 yThr..... 787
182 AAAAGAATCCGGCGGTAGTATTACTGCGCGGCTTCAGGCAAAATCGCC 231
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
788 SerArg.ArgGlyArg..... 792
232 GCTATTACCGTGGCGAAAGCGGTACTTCAGTCAGTCAGTCGATGCGGT 281
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
793 .....SerProSerProLysProArg...GlyLeuGlnArgSer...Arg 805.
282 TGAAGGCAACGACGAATCGAGTTCGACCGTACGTACCTGAAGCGGTGG 331
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
806 SerArgSerArgGluLysThrArgThrArgArg..... 818
332 CAAAAATTGAGCAGCAAAAAGTGGCGGCAACCTGATTCATCAGGCTTA 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
819 .....ArgAspArgSerGlySerSerGlnSerThrSerArgArg... 831
382 TGGACTCGGTTGCGACCCCTCCGTTTCAGCAAAATCCCTGCGGTAGATGC 431
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
832 .....ArgGlnArgSerArgSerArgSer 839
432 CGAGCCGTCGCCATCTTCGTCATGCGGTGA...CACCAATCCGCTGG 478
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
840 ArgValThrArgArgArgArgGlySerGlyThrHisSerArgSerPr 856
479 CTGCGGACCCCTACGGTCATCATCAAGAGCCGCCGA.....AGAC 519
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
856 oAlaArgGlnLysSerSerArgThrSerSerArgArgArgGlyArgS 873
520 TTCAACCGCGCTGTGTTGATTTAGCGGCTGACCGAAGCTAAATCCA 569
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
873 erArgThrProProThrSerArgLysArgSerArgSerArgThrSerPro 889
570 TGTGTGTAAGACGAGCGGCGAGCTGCGCTGGAATGCTGCCAATA 619
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
890 AlaProTriLysArgSerArgSerArgAla..... 899
620 TCGAAACACATGAATTTGGGGCCCGCATCTTCGCGGCTTGAGTGCACG 669
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
900 .....SerProAlaThr.....HisA 905
670 CACATTTCATTCATCGA.....GCCAGTCGGCGCAATAAAC 707

```

```

905 rgArgSerArgSerArgThrProLeuLysSerArgArgArgSerArgSer 921
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
708 CGTGTGGACCATCAATATCAAGACGCTGCTATCGACGCTGTTGTCG 757
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
922 ArgThrSerProValSerArgArgSerArgSerArgThrSerValTh 938
758 TAAACAGCGCTGCTGAATACCGAGCGCGTGTGCTTGGGCGGCTGCAA 807
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
938 rArgArgArgSerArgSerArgAlaSerProValSerArgArgSerA 955
808 GTCACAAACCGCGCTCTTCGCTACCGCTTTTGGGTGCGAAGGTGCTCA 857
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
955 rgSerArgThrProProValThrArgArgSerArgSerArgThrPro 971
858 ACTTACCGCGCGCAATTTGTTGACGCGGACACACCGCGGTGTCGGTT 907
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
972 ThrThrArgArgSerArgSerArgThrProPro..... 983
908 CGGTATTGAACGGTGCATTCACAGCGCGCATGATTATTGCGGACGC 957
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
984 .....ValThrArgArgSerArgSerArgSerArgThrP 994
958 TACCACATCAGATTTCCGTTATCGAAGAGCGCGCAGCAAGAGCTGT 1007
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
994 roProValThrArgArgSerArgSerArgThrSerProLleThrArg 1010
1008 CGGTGGTGTTCGCGCGCAGCGGCAAAATCTCATCAGCGCACCATC 1057
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1011 ArgArgSerArgSerArgThrSerProValThrArgArgSerArgSe 1027
1058 TCGG.....CCATTTCCCTAAAAAACAATCTTCAAGTTCAGCAGCGC 1101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1027 rArgThrSerProValThrArgArgSerArgSerArgThrSerProv 1044
1102 GTCAACGGCGCGCGCATGTCACGATCGGCACCTTATGACGCGCT 1151
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1044 alThrArgArgSerArgSerArgThrProProAlaIleArgArg 1060
1152 AATCGCGTGGACATCTCGCTACCTTGTGTCGCGATTTAATCGTGG 1201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1061 SerArgSerArgThrProLeu.....LeuProArgLy 1071
1202 GCATACCGACGAGCGCGCAGGCTTTGGTGTGCTTGAATTCAGCA...A 1248
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1071 sArgSerArgSerArgSerProLeuAlaIleArgArgSerArgSerA 1088
1249 GAAGACCTCGCTTGTGCAG.....CTTCGCTGCGCGCGG 1283
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1088 rgThrProArgThrAlaArgGlyLysArgSerLeuThrArgSerPro 1104
1284 CAAATACGAATACGCGCGCTGTCGCGAAGTGT 1319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1105 AlalleArgArgSerAlaSerGlySerSer 1116

```

seq_name: sp_human:060382

seq_documentation_block:

ID 060382 PRELIMINARY; PRT; 1791 AA.

AC 060382;

DT 01-AUG-1998 (TREMBLrel. 07, Created)

DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE KIAA0324 (FRAGMENT).

GN KIAA0324.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_taxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Ricke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,

RA Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S.,

RN SEQUENCE FROM N.A.
 RA Ohtaki S., Umeki K., Sawada Y.;
 RT "Homo sapiens mRNA for RNA binding protein, complete cds."
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB016092; BAA83718.1; -
 DR InterPro; IPR002965; P_RICH_extensions.
 DR PRINTS; PR01217; PRICHEXTENS.
 SQ SEQUENCE 2752 AA; 299672 MW; 109C64F181097123 CRC64;

alignment_scores:
 Quality: 141.00 Length: 446
 Ratio: 0.701 Gaps: 17
 Percent Similarity: 45.067 Percent Identity: 25.336

alignment_block:

US-09-303-518D-131 x Q9UQ35 ..
 Align seg 1/1 to: Q9UQ35 from: 1 to: 2752
 32 CCATCGCGGCGACGCGAGCAAGTCATTATGACGCGCGCGCATACC 81
 1721 ProProArgHisArgArgSerProSerValSerSerProGluProAlaG1 1737
 82 GAAGTCGCGTGTGGCGAAGAAATATGCGCATCGCGCCCTCGATGAA 131
 1737 uLysSerArgSerArgArgArgArgSerAlaSerProArgThrL 1754
 132 AATCAAGGAGGTGAGCGCTCAAAAAGCCCAAGTGTGTTGAAGACA 181
 1754 yThr.....Thr 1756
 182 AAAAAGATCGGCGAGTAGTATTACTGCGCGCGCTTCAGGCAAAATCGCC 231
 1757 SerArg, ArgGlyArg..... 1761
 232 GCTATTACCGTGGCGAAGCGCGTACTGCTAGTCGTGATTCGGT 281
 1762SerProSerProLysProArg...GlyLeuGlnArgSer...Arg 1774
 282 TGAAGCAACGAGAAATCGAGTTCGAACGCTACGTACCTGAGCGCTGG 331
 1775 SerArgSerArgGlnLysThrArgThrArgArg..... 1787
 332 CAAAAATGACGAGCAAAAGTGCAGCGCAACCTGATTCAATCAGCGCTTA 381
 1788ArgAspArgSerGlySerSerGlnSerThrSerArgArg.... 1800
 382 TGGACTGCGCTTCGCAACCGCTCGGTTCAGCAAAATCCCTGCGGTAGATGC 431
 1801ArgGlnArgSerArgSerArgSer 1808
 432 CGAGCGGTTCGCATCTTCGTCATCGATGCA...CACCAATCGCGTGG 478
 1809 ArgValThrArgArgArgArgGlySerGlyTyrHisSerArgSerPr 1825
 479 CTGCGGACCTACGTCATCATCAAGAGCGCGCA.....AGAC 519
 1825 oAlaArgGlnGluSerSerArgThrSerSerArgArgArgArgGlyArgS 1842
 520 TTCAACGCGGCTGTGTGTTATGAGCCCTGACCGAAGCTAAATCCA 569
 1842 erArgThrProProThrSerArgLysArgSerArgSerArgThrSerPro 1858
 570 TGTGTGTAAGCAGCAGCGCGAGCTGCGCTGTAATGTCGCCAATA 619
 1859 AlaProTrpLysArgSerArgSerArgAla..... 1868
 620 TCGAAACACATGAATTTGCGCGCGGCAATCCCTGCGGCTTGTAGTGCACG 669
 1869SerProAlaThr.....HisA 1874

670 CACATTCATTTTCATCGA.....GCCAGTCGCGCGCAATAAAAC 707
 1874 rgArgSerArgSerArgThrProLeuLysSerArgArgSerArgSer 1890
 708 CGTGTGGACCATCAATTCAGACGCTGATTCGATCGGACGTTTGTTCG 757
 1891 ArgThrSerProValSerArgArgSerArgSerArgThrSerValTh 1907
 758 TAACAGCGCGTCTGAATACCGAGCGCTGTTGCTTGGCGCGCTGCAA 807
 1907 rArgArgArgSerArgSerArgAlaSerProValSerArgArgSerA 1924
 808 GTCAACAAACCGCGCTCTTCGTCACCTGTTGGTGCAGAGGTGCTCA 857
 1924 rgSerArgThrProValThrArgArgSerArgSerArgSerArgThrPro 1940
 858 ACTTACCGCGCGGAATTTGTTGACGCGGACACCGCTGATTTCCGGTT 907
 1941 ThrThrArgArgSerArgSerArgThrProPro..... 1952
 908 CGGTATTGAACGCGTCCGATTGCACAAGCGCGCATGATTATTGGGACGC 957
 1953ValThrArgArgSerArgSerArgSerArgThrP 1963
 958 TACCACAATCAGATTTCCGTTATCGAAGAGCGCGCAGCAAGAGCTGT 1007
 1963 roProValThrArgArgSerArgSerArgThrSerProLleThrArg 1979
 1008 CGCTGCGTGGCGGCGAGCGGCAATACTCCATCACGCGCACCACTC 1057
 1980 ArgArgSerArgSerArgThrSerProValThrArgArgSerArgSe 1996
 1058 TCGG.....CCATTCTCTAAACCAACTCTTCAAGTTCACGACAGCC 1101
 1996 rArgThrSerProValThrArgArgSerArgSerArgSerArgThrSerProv 2013
 1102 CTCACGCGCGGCGAGCGCGCATGTCACGATCGCGCTTATGAGCGCT 1151
 2013 alThrArgArgSerArgSerArgThrProProAlaLleArgArgArg 2029
 1152 AATGCGGTGGACATCTGCTACCTGCTTGTTCGCGGATTAATCGTCG 1201
 2030 SerArgSerArgThrProLeu.....LeuProArgLy 2040
 1202 CGGTATACGACGAGCGCGCGCTTGGTGTGCTTGAATTCGACGA...A 1248
 2040 sArgSerArgSerArgSerProLeuAlaLleArgArgSerArgSerA 2057
 1249 GAAGACCTCGCTTTGTGCGAG.....CTTCGCTCTCCCGGG 1283
 2057 rgThrProArgThrAlaArgGlyLysArgSerLeuThrArgSerProPro 2073
 1284 CAATACCAATACGCGCGCTGTTGCGCAAGTGTCT 1319
 2074 AlaLleArgArgSerAlaSerGlySerSerSer 2085
 seq_name: sp_bacteriap:Q9WY86
 seq_documentation_block:
 ID Q9WY86 PRELIMINARY; PRT; 451 AA.
 AC Q9WY86;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE ELECTRON TRANSPORT COMPLEX PROTEIN, PUTATIVE.
 GN TM0244.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogales; Thermotoga.
 OX NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB8 / DSM 3109;
 RX MEDLINE=99287316; PubMed=10360571;

101TCGCAAGCAGCGAGCTTCGGTAATGGCGG 70
 788 ProTyAspIleValTyrSerProSerThrPheAlaAlaIleSerSerG1 804
 69 G 69
 804 Y 804

seq_name: sp_human:Q9UHA8

seq_documentation_block:

AC Q9UHA8 PRELIMINARY; PRT; 2296 AA.

DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE SPLICING COACTIVATOR SUBUNIT SRM300.
 GN SRM300.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]

SEQUENCE FROM N.A.

RP MEDLINE=20132238; PubMed=10668804;

RA Blencowe B.J., Bauren G., Eldridge A.G., Issner R., Nickerson J.A.,

RA Rosonina E., Sharp P.A.;

RT "The SRM160/300 splicing coactivator subunits.";

RL RNA 6:111-120(2000).

DR EMBL; AF201422; AAF21439.1; -.

DR InterPro; IPR002965; P-rich_extensn.

DR PRINTS; PR01217; PRICHEXTENS.

SQ SEQUENCE 2296 AA; 251964 MW; 17C0BD4EA10A9CF9 CRC64;

alignment_scores:

Quality: 127.00 Length: 418
 Ratio: 0.722 Gaps: 17
 Percent similarity: 42.105 Percent identity: 24.641

alignment_block:

US-09-303-518D-131 x Q9UHA8 ..

Align seg 1/1 to: Q9UHA8 from: 1 to: 2296

32 CCATCGGGCGGCGAGCGAGTCAATTTATGACGGCGCGGCGCATACC 81
 1721 ProProhghisArgArgSerProSerValSerSerProGluProAlaG1 1737
 82 GAAGTCGCGTGTGGCGGAAGATATGTCGGCATGCGCCCTCGATGAA 131
 1737 uLysSerArgSerSerArgArgArgSerAlaSerSerProArgThrL 1754
 132 AATCAAGGAGGTGAAGCCGTCAAAAAGCCCAAGTCTGCTTTGAAGACA 181
 1754 yThr.....
 182 AAAAGAAATCCGGCGTAGTATTACTGCGCGCGCTTCAGGCAAAATCGCC 231
 1757 SerArg. ArgGlyArg..... 1761
 232 GCTATTCACGCTGGCGGAAAGCGGTACTTTCAGTCAGTCGTGATGCCGT 281
 1762SerProSerProLysProArg...GlyLeuGlnArgSer...Arg 1774
 282 TGAAGCAACGACGAATTCAGTTCGACGCTACGTACCTGAAGCGCTGG 331
 1775 SerArgSerArgGluLysThrArgThrArg..... 1787
 332 CAAATTTGACGAGGAAAGTGGCGCAACCTGATTCAATCAGGCTTA 381
 1788ArgAspArgSerGlySerSerGlnSerThrSerArgArg.... 1800

382 TGGACTCGGCTTCGACCCCGCTCCGTTTCAGCAAAATCCCTGCCGTAGATGC 431
 1801ArgGlnArgSerArgSerArgSer 1808
 432 CGAGCCGTTGCCATCTTCGTCATCGCATGGA...CACCAATCCGCTGG 478
 1809 ArgValThrArgArgArgGlyGlySerGlyTyrHisSerArgSerPr 1825
 479 CTGCGGACCCCTACGTCATCATCAAGAAGCGCGCA.....AGAC 519
 1825 oAlaArgGlnGluSerSerArgThrSerSerArgArgArgGlyArgS 1842
 520 TTCAAGCGCGCGCTTTGATTGAGCCGCTGACCGAAGCTAAATCCA 569
 1842 erArgThrProProThrSerArgLysArgSerArgSerArgThrSerPro 1858
 570 TGTGTGTAACGACGAGCGCGCAGCGTCCGCTCTGAAATGCTGCCAATA 619
 1859 AlaProTrpLysArgSerArgSerArgAla..... 1868
 620 TCGAAACACATGAATTTGGCGCGCGCATCTCGCGGCTTCAGTGGCAGC 669
 1869SerProAlaThr.....HisA 1874
 670 CACATTCATTTCATCGA.....GCCAGTCGGCGGGAATAAAC 707
 1874 rgArgSerArgSerArgThrProLeuIleSerArgArgSerArgSer 1890
 708 CGTGTGGACCATCAATATCAAGACGTGATGCTATCGGAGTTTGTTCG 757
 1891 ArgThrSerProValSerArgArgSerArgSerArgThrSerValTh 1907
 758 TAACAGGCGCTGTAATACCGAGCGGTGCTTGGCGCGGCGCTCAA 807
 1907 rArgArgArgSerArgSerArgAlaSerProValSerArgArgSerA 1924
 808 GTCAACAAACCGCGCTCTTCGTCACGTTTGGTGGCGAAGGTGTCTCA 857
 1924 rgSerArgThrProProValThrArgArgSerArgSerArgThrPro 1940
 858 ACTTACCGCGCGCGAATTTGTTGACCGGCAACCGGCTGATTTCCGGTT 907
 1941 ThrThrArgArgArgSerArgSerArgThrPro..... 1952
 908 CGGTATTGAACGCTCGGATTCACAAAGCGCGCATGATTATTGGGACGC 957
 1953ValThrArgArgSerArgSerArgSerArgThrp 1963
 958 TACCACAATCAGATTCGTTATCGAAGAGCGCGCACGAAGAGCTGTT 1007
 1963 roProValThr.....ArgArgArgSerArgSerArg..... 1973
 1008 CGGCTGGGTTGGCGCGCAGCGGCAATACTCCATCATCAGCGCACCATC 1057
 1974Th 1974
 1058 TCGGCGCATTTCTAAAAACAACCTTCAGTTTCACGACAGCGCTCAAC 1107
 1974 rSerProIleThrArgArgSerArgSerArgThrSerProValThrA 1991
 1108 GCGCGGCGCGCGCATGTTACCGATCGCCTTATGAGCGCGTAATGCC 1157
 1991 rgArgArgSerArgSerArgThr..... 1998
 1158 GTTGACATCCTGCTACCTGCTTTTCGCGGATTTAATCGTCGCGGATA 1207
 1999SerProValThrArgArgArgSe 2006
 1208 CGCAGCGCGCGCAGCTTTGGGTTGCTTGGAAATTGGAGAGAGACCTC 1257
 2006 rArgSerArgThrSer.....ProValThrArgArgSerA 2019
 1258 GC 1259


```

91  ACAGCACTTCG 81
   ::::|||||
826  erSerSerSer 829
seq_name: sp_rodent:Q99K60

seq_documentation_block:
ID  Q99K60      PRELIMINARY;      PRT;      631  AA.
AC  Q99K60;
DT  01-JUN-2001 (TrEMBLrel. 17, Created)
DT  01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT  01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE  MUCIN 1, TRANSMEMBRANE.
GN  MUC1.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_taxid=10090;
RX  [1]
RN  SEQUENCE FROM N.A.
RP  Strausberg R.;
RL  Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR  EMBL; BC005441; AH05441.1; -.
DR  MGD; MGI:97231; Muc1.
DR  InterPro; IPR000082; SEA.
DR  Pfam; PF01390; SEA; 1.
DR  SMART; SM00200; SEA; 1.
DR  PROSITE; PS50024; SEA; 1.
KW  Transmembrane.
SQ  SEQUENCE 631 AA; 64690 MW; B4236DF6655F4C3 CRC64;

alignment_scores:
  Quality: 122.00      Length: 467
  Ratio: 0.604        Gaps: 22
Percent Similarity: 43.255 Percent Identity: 23.769

alignment_block:
US-09-303-518D-131/rev x Q99K60
Align seg 1/1 to: Q99K60 from: 1 to: 631

1337 TCCTTCTCAATGTTTCCAGCACTTCCGCAACAGCGCGCGTATTCGTA 1288
   ||| ||:::|||||:::|||||
38  SerSerLeuAlaSerThr..... 45

1287 TTTCCCGGCGAGAGCTGCAACAGCGAGGTCTTCTTCGTCCAAAT 1238
   ||| ||:::|||||:::|||||
46  .....ThrProValHisSer.....SerAsnSerAspProA 56

1237 CCAAGCAACCAAGCCTGCGCGCTGTCGGTATCGCGCGAGATTAAATCG 1188
   ::::|||||::: ||| ||:::|||||:::|||||
56  laThrArgProGlyAspSerThrSer...SerProValGlnSerSer 71

1187 CGCAAAAGCAAGGTAGGAGATGTCACAGCGGATTAACGCGCTCATAGT 1138
   ::::|||||::: ||| :::: ||| :::: |||
72  ThrSerProAlaThrArgAlaProGluAspSerThrSerThrAlaVa 88

1137 GCCGATCGGTACCATGCGCGGCTGCGCGGTTGACGCGTGTGCTGAAT 1088
   | ||||| ||||| ||||| |||||
88  lLeuSerGlyThr.....SerSerProAlaThrThrAlaProValAsnS 103

1087 TGAAGAGTGTGTTTTAGGAAATGGCGGAGAGTGGTGGCGGTGATGGAG 1038
   ||| |||
103  erAlaSer.....Ser 106

1037 TATTGTCCGCGCTGCGCGCAACCCGCAACAGCACTTTTGTGCGGCC 988
   ::::|||||::: ||| :::: ||| :::: |||
107  ProValAlaHisGlyAspThrSerSerProAlaThrSerLeuSerLysAs 123

987 TTCTTCGATACGGAAATCTGATGTGGTAGCGTCCCAATAATCATGCG 938
   ||||| :::: ||| :::: |||

```

```

181 TGCTTCAACAGCACTGGCCCTTTTGTGACGGCTTCACCTCTCTGATT 132
|||||:||||| ||| :|||
365 ..ValSerAsnGlyThrGlnPro.....SerValProSerGlnTyr 377
|||||:||||| ||| :|||
131 TTCATCCAGGGCGCATCCGACATATCTCCCAAGCAACCGCACTTC 82
|||||:||||| ||| :|||
378 ProValSerProThrMetAlaThrThrSerSerHisSerThrIleAla 394
|||||:||||| ||| :|||
81 G 81
394 r 394

```

seq_name: sp_human:Q9UF83

```

seq_documentation_block:
ID Q9UF83 PRELIMINARY; PRT; 580 AA.
AC Q9UF83;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 61.1 KDA PROTEIN (FRAGMENT).
GN DFEZP434C196.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE-TESTIS;
RC Koehler K., Beyer A., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL133561; CAB63715.1; -.
DR InterPro; IPR002965; P_rich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 580 AA; 61089 MW; C46E82F050F06547 CRC64;

```

```

alignment_scores:
Quality: 121.50 Length: 530
Ratio: 0.557 Gaps: 23
Percent Similarity: 41.132 Percent Identity: 23.019

```

alignment_block:

US-09-303-518D-131/rev x Q9UF83 ..

Align seg 1/1 to: Q9UF83 from: 1 to: 580

```

1336 CCTTCTCAATGGTTTCCAGCACTTCCGCAACAGCGGGCGGTATTCGTAT 1287
|||||:||||| ||| :|||
29 ProSerArgAlaSerProThrArgMetProSerArgAlaSerLeuLysMe 45
|||||:||||| ||| :|||
1286 TTGCGCGGCGCAGCAAGCTGCACAAAG.....CGAGTCTCTCTTC 1246
|||||:||||| ||| :|||
45 tthrProPheArgAlaSerLeuThrLysMetGluSerThrAlaLeuLeu 62
|||||:||||| ||| :|||
1245 GTCCCAATCCCA..... 1235
|| :|||
62 rgThrLeuProArgAlaSerLeuMetArgThrProThrArgAlaSerLeu 78
|||||:||||| ||| :|||
1234 .....AGCAA 1230
79 MetArgThrProProArgAlaSerProThrArgLysProProArgAlaSe 95
|||||:||||| ||| :|||
1229 CCCAAG...CTGCGCGCTGTCGGTATCGCGCAGGATTAAATCGCGCAA 1183
|||||:||||| ||| :|||
95 rProArgThrProSerArgAlaSerProThrArgArgLeuProArgAla 112
|||||:||||| ||| :|||
1182 AGCAGAGTAGGCAGGATGTCACAGCGCATTTACGGCTCATTAAGTGC 1133
|||||:||||| ||| :|||
112 erPro.....MetGlySerProHis.....ArgAlaSerProMetArg 124

```

```

1132 TCGTAGCATGCGCGGTGCGCGCGTGTGACGGCTGTCGTAACCTGAAG 1083
|||||:||||| ||| :|||
125 ThrProPro.....ArgAlaSerProThr..... 132
|||||:||||| ||| :|||
1082 AGTTTGTTTTGTAGGAATGGCGGAGAGTGGTGGCGGTGATGGAGTATT 1033
|||||:||||| ||| :|||
133 .....GlyThrProSerThrAlaSerProThrGlyT 143
|||||:||||| ||| :|||
1032 GTCCGGCTGCGCGCAACCCGACCGCAAGCTCTTTGCTGGCGCTTCTT 983
|||||:||||| ||| :|||
143 hrProSerSerAlaSerProThrGlyThrProPro..... 154
|||||:||||| ||| :|||
982 CGATAACGGGAATCTGATTGGTAGCTGCCCAATATCATGCGCGCT 933
|||||:||||| ||| :|||
155 .....ArgAlaSerProThrGlyThrProProArgAlaThrAr 169
|||||:||||| ||| :|||
932 TGTCAATCGCACCGCTTCAATA.....CCGAACCGGAATCA.... 896
|||||:||||| ||| :|||
169 gSerProSerThrAlaSerLeuThrArgThrProSerArgAlaSerLeu 186
|||||:||||| ||| :|||
895 .....CGCGGTGT 887
|||||:||||| ||| :|||
186 hrArgTirProProArgAlaSerProThrArgThrProProArgGluSer 202
|||||:||||| ||| :|||
886 CGCGCTCAACCAATTCGCGCGGTAACTTGCAGACACCTTC....GCACC 841
|||||:||||| ||| :|||
203 ProArgMetSerHisArgAlaSerProThrArgThrProProArgAlaSe 219
|||||:||||| ||| :|||
840 CAAACAGGTAGCAGAGCGCGGTTTGTGACTTGCAGCGCGCCCAAGG 791
|||||:||||| ||| :|||
219 rProThrArgArgProProProArgAlaSerProThrArgThrProProArg 236
|||||:||||| ||| :|||
790 CAACACCGCTCGGTATTCAGAGCGGCTGTACGAACAACAGTCCGATA 741
|||||:||||| ||| :|||
236 luSerLeuArgThrSerHisArgAlaSerProThrArgMetProProArg 252
|||||:||||| ||| :|||
740 GCAATCAGCTCTGATAATGATGTCACACACGGTTTATTTCGCGCGCAG 691
|||||:||||| ||| :|||
253 AlaSerProThr.....ArgArgProProArgAlaSerProTh 265
|||||:||||| ||| :|||
690 TGGCTCGATGAATGATGTCGCGCTCACTCAACGCGGAGGATGCGGG. 642
|||||:||||| ||| :|||
265 rGlySerProPro...ArgAlaSerProMetThrProProArgAlaSerP 281
|||||:||||| ||| :|||
641 .....CGCCCAATTCATGT 627
|||||:||||| ||| :|||
281 rArgThrProProArgAlaSerProThrThrProSerArgAlaSer 297
|||||:||||| ||| :|||
626 GTTTCGATATTGGCAGCATTTTCAGCGGACGCTGCGCTGCTGCTTT 577
|||||:||||| ||| :|||
298 LeuThrArgThrProSerTipAlaSerProThrThrThrProSerArgAl 314
|||||:||||| ||| :|||
576 ACACACATGGATTTAGCTTCGGTCAGCGGCTCAATACCAACAGCGCGC 527
|||||:||||| ||| :|||
314 aSerLeuMetLysMetGluSerThrValSerIleThrArgThrProPro 331
|||||:||||| ||| :|||
526 GTTTCGAG.....TCT 516
|| :|||
331 rgAlaSerProThrGlyThrProSerArgAlaSerProThrGlyThrPro 347
|||||:||||| ||| :|||
515 TCGCGGGTCTTTGATGATGACCGTAGGTCGGCAGCGCGGATGGT 466
|||||:||||| ||| :|||
348 SerArgAlaSerLeu.....ThrGlySerProSerArgAlaSerLeuTh 362
|||||:||||| ||| :|||
465 GTCCATCGATTCAGCAAGATGGCGGACGCTGGCATCTACGCCAGGA 416
|||||:||||| ||| :|||
362 rGlyThrProSerArgAlaSerLeuIleGlyThrProSerArgAlaSerL 379
|||||:||||| ||| :|||
415 TTTTGTCTGAACGGAGCGGTGCGAAGCGCAGTCCCATAGCCTGATTGAATC 366
|||||:||||| ||| :|||
379 euIleGlyThrProSerArgAlaSerLeuThrGlyThrProProArgAla 395

```



```

395 CGAAGCCAGCTCCATAAGCCTGATTGAATCAGGTTCGGCGCACATTTTC 346
||||| :|||
277 SerSerSerSerPro.....ThrSerThrSerThrIleSe.291.
||||| :|||
345 GCTGCTCAATTTCCAGCGCTTCAGGTACGTAGCTGCGTTCGAATCGATTT 296
| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
291 rSerSerSerSerSerSerProThrSerThrSerThrIleSe.308
295 CGTCGTTGCTTCAACGGCAATCAGCACTGACTGAAGTACGCGCTTTTCG 246
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
308 erSerSerSerSer.....SerSerSerPheSer 318
245 CCACGGTGAATACGGCGATTTTCCTGAGCGCGCGCACTAATACTAC 196
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
319 SerThrLeuSerSerSerMetSerSerSerSerPheSerSe.335
195 GCCCGGATCTTTTCTCAACAGCACTTGGCTTTTTCAGCGCTT 146
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
335 rPro.....ThrSerSerSerThrIleSerSerSerSerSe.349
145 CACCTTCCTGATTTTCATCGAGGGCGCATGCGGACATATTTCTCGCCA 96
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
349 erProSerSerSerSerPheSerSerThrThrSerSerSerIysSerSer 365
95 AGCAACCGGACTCGGTATGCGCGGCGCTCATAAATGACTGCTCCGG 46
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
366 SerSerPheSerThrValSerSerSerSerSerThrSerSerSerTh 382
45 TCTG 42
|||
382 rLeu 383

```

seq_name: sp_human:Q9UQ39

```

seq_documentation_block:
ID Q9UQ39 PRELIMINARY; PRT; 956 AA.
AC Q9UQ39;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE RNA BINDING PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtaki S., Umeki K., Sawada Y.;
RT "Homo sapiens mRNA for RNA binding protein, partial cds.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016088; BAA83714.1; -.
DR InterPro; IPR002965; P_Rich_extensn.
DR PRINTS; PR01217; PRICEXTENS.
FT NON_TER 1
FT NON_TER 956
SQ SEQUENCE 956 AA; 106462 MW; C9FEAD07B49F5011 CRC64;

```

```

alignment_scores:
Quality: 118.50 Length: 398
Ratio: 0.598 Gaps: 19
Percent Similarity: 49.749 Percent Identity: 25.126

```

alignment_block:

US-09-303-518D-131 x Q9UQ39

Align seq 1/1 to: Q9UQ39 from: 1 to: 956

```

18 AGTCTAAATCTGCCCATCGCGGCGAGACCGGAGCAAGTCATTATGACG 67
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
518 ArgSerArgSerAlaArgGlyArg.....SerHisSerArgSe 531
68 GCCC.....GGCCATTACCGAGTCCGCTTGTGGCGGAAGATAT 108

```

```

||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
531 rProAlaThrArgGlyArgSerArgSerArgThrProAlaArg..... 546
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
109 GTCCGATCGCCCTCGATGAAATCAAGGAAGGTGAAGCCGTCACAAA 158
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
547 .....GlyArg...SerArgSerArg 552
159 AGGCCAAGTCTGTTGAAGACAAAGAACATCCGGCGCTAGTATTACTG 208
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
553 ThrProAla.....ArgArgArgSerArgSerArgThr..... 563
209 CGCCGGCTTCAGGCAAAATCGCGCTATTACCGGTGGCGAAGCGCGTA 258
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
564 .ProThrArgArgArgSerArgSerArgThrProAlaArgArgGlyArg. 579
259 CTTAGTCAGTCGTGATTGCCGCTTGAAGGCAACAGCAATCGAGTTGCA 308
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
580 .....SerArgSerArgThrProAlaArgArgSerArgThrArg 593
309 ACG.....CTACGTACCTGAAGCGCTGGCAAAATTTGACGACGCAAAAGTCC 355
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
594 SerProValArgArgArgSerArgSerArgSerProAlaArgArgSerGI 610
356 GCCGCAACCTGATTCATCAGGCTTATGGACTTCGCGCTTCGCAACCGCTCG 405
|||
610 Y.....ArgSera 613
406 TTCACAAAATCCCTGCGTAGATCCGCGGTTCGCCATCTTCGTCAA 455
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
613 rgSerArgThrProAlaArgGlyArgSer...ArgSerArgThrPro 628
456 TCGCATGGACACCAATCGCTGGCTGCGACCTACGGTCAATCATCAAG 505
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
629 AlaArgArgGlyArgSerArgSerArgThrProAlaArgArgSerGlyAr 645
506 AAGCCGCGCAAGACTTCAACCGCGCTTGTGGTATTGAGCGCGCTGACC 555
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
645 gSerArgSerArgThrProAlaArg.....ArgGlyArgSera 658
556 GNACTAAATCCATGTGTAAAGCAGCAGCGCGCGCTGCA 605
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
658 rgSerArgThrProArgArgGlyArgSerArgSerArgSerLeuValArg 674
606 AAATGCTGCCAATATCGAAACACATGATTTGGCGCGCCGCTGTCGCG 655
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
675 Arg.....GlyArgSerHisSerArgTh 682
656 GCTTGAGTGGCAGCACATTCATTCATCGAGCGCGCGCAATAAA 705
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
682 rProGlnArgArgGlyArgSerGlySerSerGluArgLysAsnLysS 699
706 ACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACCTTGT 755
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
699 erArgThrSerGlnArgArgSerArgSerArgSerProGluMetLys 715
756 CGTACAGCGCGTCTGAATACGAGCGGTGCTTGGCTTGGCGCGCGCTGC 805
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
716 LysSerArgIleSerArgArgSerArgSer...LeuSerSerProAr 731
806 AAGTCAACAAACCGCGCTCTTGGCTACCGCTTTTGGGTGCGAAGTGTCT 855
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
731 gSerLysAlaLysSerArgLeuSerLeuArgSerLeuSerGlySerS 748
856 CAACTTACCGCGCGAATTGTTGACGCGGACAAACCGCGGTGATTTCGG 905
||| :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
748 erProCysProLysGlnLysSer...GlnThrProProArgArgSerArg 763
906 TTCGTATTGAACGGTGGATTGCAAGCGCGCATGATTATTATTTGGGAC 955
|||
764 SerGly.....SerSe 767
956 GCTACCAACATCAGATTTCGTTATCGAAGAGCGCGCAAGAGCTG 1005
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```



```

767 rGlnProLysAlaLysSerArgThrProProArg.ArgSer.ArgSerSe 783
1006 TTGGCTGGTTCGGCGCGAGCGGCAATAATCTCATCAGCGGCACCAC 1055
||||| : : : : : : : : : : : : : : : : : : : : : : : :
783 rSerSerProProLysGlnLysSerLysThrProSerArg..... 797
1056 TCTCGGCATTCTCTAAACAACTCTTCAAGTTTCAGACAGCGTCA 1105
798 .....GlnSerHisSerSerSerProHisProLys 808
1106 ACGCGCGCGACCGCATGGTACCGACATTCATGAGCGCGTAATG 1155
ValLysSerGlyThrProProArgGlnGlySerIleThrSerProGlnAl 825
1156 CGGTGGACATCTCGCTACCT.....TCCTTT 1183
: : : : : ||| |||||
825 aAsnGlnSerValThrProGlnArgArgSerCysPhe 838

seq_name: sp_human:Q9UQ40
seq_documentation_block:
ID Q9UQ40 PRELIMINARY; PRT; 1262 AA.
AC Q9UQ40;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE RNA BINDING PROTEIN (FRAGMENT).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtaki S., Umeki K., Sawada Y.;
RT "Homo sapiens mRNA for RNA binding protein, partial cds.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016087; BAA83713.1; -
DR InterPro; IPR002965; P.Rich.extensin.
DR PRINTS; PR01217; PRICEXTENSIN.
FT NON_TER 1
FT NON_TER 1262
SQ SEQUENCE 1262 AA; 138003 MW; 75E76BD73FA14295 CRC64;

alignment_scores:
Quality: 118.50 Length: 398
Ratio: 0.598 Gaps: 19
Percent Similarity: 49.749 Percent Identity: 25.126

alignment_block:
US-09-303-518D-131 x Q9UQ40
Align seg 1/1 to: Q9UQ40 from: 1 to: 1262
18 AGGTCTAAATCTGCCATCGCGGCGAGCGGAGCAAGTCATTATTGACG 67
||||| : : : : : : : : : : : : : : : : : : : : : : : :
343 ArgSerArgSerAlaArgArgGlyArg.....SerHisSerArgSe 356
68 GCCC.....GCCATTACCGAAGTCGCTTGGCGGAAGATAT 108
||| : : : : : : : : : : : : : : : : : : : : : : : :
356 rProAlaThrGlyArgSerArgSerArgThrProAlaArg..... 371
109 GTCGGCATGCGCCCTCGCTCGATGAAATCAAGGAAGGTGAAGCGGTCAAAA 158
||||| : : : : : .....GlyArg.....SerArgSerArg 377
372 .....
159 AGGCCAAGTCTGTTTGAAGACAAAAGATCCGGCGTAGTATTACTG 208
||||| : : : : : : : : : : : : : : : : : : : : : : : :
378 ThrProAla.....ArgArgSerArgSerArgThr..... 388
209 CGCGGCTTACGCAAAATCCCGCTATTTCACCGTCGCGAAAGCGGTA 258
||||| : : : : : : : : : : : : : : : : : : : : : : : :
389 .ProThrArgArgArgSerArgSerArgThrProAlaArgGlyArg. 404

```

```

259 CTTCACTCAGTCGTGATTCCGTTGAAGGCAACGACGAAATCGAGTTCGA 308
||||| : : : : : : : : : : : : : : : : : : : : : : : :
405 .....SerArgSerArgThrProAlaArgArgArgSerArgThrArg 418
309 ACG...CTACCTACCTGAAGCGCTGGCAAAATTCAGCAGCGCAAAAGTGC 355
: : : : : ||| : : : : : ||||| : : : : : : : : : : : :
419 SerProValArgArgSerArgSerArgSerProAlaArgArgSerG1 435
356 GCCGCAACCTGATTCAATCAGGCTTAAGGACTGCGCTTCGACCCGCTCG 405
435 y.....ArgSerA 438
406 TTCAGCAAAATCCCTGCGTAGATCCGAGCGCTTCGCCATCTTCGTCAA 455
: : : : : ||||| : : : : : ||||| : : : : : : : : : : : :
438 rSerArgThrProAlaArgArgGlyArgSer...ArgSerArgThrPro 453
456 TCGCATGGACACCAATCCGCTGGTGGCGACCCCTACGGTTCATCATCAAG 505
: : : : : ||||| : : : : : ||||| : : : : : : : : : : : :
454 AlaArgArgGlyArgSerArgSerArgThrProAlaArgArgSerGlyAr 470
506 AAGCCGCGAAGACTTCAACGCGGCTGTGGTATTGAGCGCGCTGACC 555
||||| ||| : : : : : : : : : : : : : : : : : : : : : :
470 gSerArgSerArgThrProAlaArg.....ArgGlyArgSerA 483
556 GAACGTAAATCCATGTGTAAAGCAGCGCGCAGACGTCGCGTCTGA 605
||| : : : : : ||||| ||||| ||||| : : : : : : : : : : : :
483 rSerArgThrProArgArgGlyArgSerArgSerArgSerLeuValArg 499
606 AAATCTGCCAATATCGAAACACATGAATTTGGGGCGCCGATCTCTGCG 655
: : : : : ||||| : : : : : ||||| : : : : : : : : : : : :
500 Arg.....GlyArgSerHisSerArgTh 507
656 GTTCAGTGGCAGCACACATTCATTCAGCGCGCGCGCAATAAA 705
: : : : : ||||| : : : : : ||||| : : : : : : : : : : : :
507 rProGlnArgArgGlyArgSerGlySerSerSerGlnArgLysAsnLysS 524
706 ACCGTGTGGACCATCAATATCAAGACGTCGATTCGTCATCGGACGTTGT 755
||||| : : : : : ||||| : : : : : ||||| : : : : : : : : : :
524 erArgThrSerGlnArgArgSerArgSerArgSerProGluMetLys 540
756 CTAACAGCGCGTCTGAATACCGCGCGTGGTTCCTTGGCGCGCGCTGC 805
||||| : : : : : ||||| : : : : : ||||| : : : : : : : : : :
541 LysSerArgIleSerArgArgSerArgSer...LeuSerSerProAr 556
806 AAGTCAACAAACCGCGCTCTTTCGTCACCGTTTGGTGCGAAGGTCTCT 855
||||| : : : : : ||||| : : : : : ||||| : : : : : : : : : :
556 gSerLysAlaLysSerArgLeuSerLeuArgArgSerLeuSerGlySerS 573
856 CAACCTACCGCGCGAATTTGGTTGACGCGGACACCGCGTGTTCGGG 905
|| : : : : : : : : : : : ||||| : : : : : ||||| : : : : :
573 erProCysProLysGlnLysSer...GlnThrProProArgArgSerArg 588
906 TTCGGTATTGACGTCGTCATTCACAGCGCGCATGATTATTGGGAC 955
||| : : : : : : : : : : : : : : : : : : : : : : : :
589 SerGly.....SerSe 592
956 GTACACCAATCAGATTTCCGTTATCAGAGAGCGCGCAGCAAGAGCTG 1005
: ||||| : : : : : ||||| : : : : : ||||| : : : : : : : : : :
592 rGlnProLysAlaLysSerArgThrProProArg.ArgSer.ArgSerSe 608
1006 TTCGGCTGGTTCGCGCGCGCGGACAAATCTCCATCAGCGCGCACCCAC 1055
||||| : : : : : : : : : : : ||||| : : : : : ||||| : : : : :
608 rSerSerProProLysGlnLysSerLysThrProSerArg..... 622
1056 TCTCGGCATTTCTCTAAACAACTCTTCAAGTTCACGACGCGCTCA 1105
: : : : : ||||| : : : : : ||||| : : : : : : : : : : : :
623 .....GlnSerHisSerSerSerProHisProLys 633
1106 ACGCGGCGCGCGCGCATGTGTACCGTACCGGACCTATGAGCGCGTAATG 1155
: : : : : ||||| : : : : : ||||| : : : : : : : : : : : :
634 ValLysSerGlyThrProProArgGlnGlySerIleThrSerProGlnAl 650

```

423 GCGAGGATTTGCTGAACGACGGTGC GAAGCGCAG 386
 :
 411 rThrProIleProThrThrSerValGlnProSerSerSerServ 428
 385 TCCATAACGCTGATGATCAGGTGGCGGCACATTTTTCGCTGCTCAAT 336
 :
 428 alProThrSerAlaThrSerValGlnThrSerSerSerThr 444
 335 TTTGCCAGCGCTTCAGGTACGTAGCTTCGAACTCGATTTTCGTGTC 286
 :
 445 ProIleProSerThrThrSerValGlnProSerSerSerSerAlaPr 461
 285 TTCACGCAATCAGCT...GACTGAAGTACGCGCTTTTCGCCACGGT 239
 :
 461 oThrThrSerAlaThrSerValGlnProSerSerSerPro.... 476
 238 GAATAGCGGCGATTTGCTGAACGCGCGCAGTAATATCTACGCCCGGA 189
 :
 477ProfileSerThrIleSerValGlnPro... 486
 188 TTCTTTTGTCTTCAACAGCACTTGCGCTTTTGTGACGGCTTCACCTTC 139
 :
 487SerSerSerSerSerProThrThrThrSerThrServa 500
 138 CTGATTTTCATCGAGGGCGCATGCCGACATATCTTTCGCCAAGCAACG 89
 :
 500 lGlnProSerSerGlySerAlaProThrThrSerAlaThrSerValG 517
 88 CGACTTCGGTAATGCGCGGCGG 66
 :
 517 lnProSerSerSerSerPro 524

seq_name: sp_plant:Q9LWY3

seq_documentation_block:

ID Q9LWY3 PRELIMINARY; PRT; 739 AA.

AC Q9LWY3;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE HYPOTHETICAL PROTEIN.

OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehrhartoideae; Oryzeae; Oryza.

OX NCBI_TaxID=4530;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. NIPPONBARE;

RA Sasaki T., Matsumoto T., Yamamoto K.;

RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC

RT clone:P0644B06."

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AP001129; BAA90631.1; .

KW Hypothetical protein.

SQ SEQUENCE 739 AA; 82154 MW; 5DC56AF5DD1CA88 CRC64;

alignment_scores:

Quality: 116.50

Ratio: 0.626

Percent Similarity: 42.661

Percent Identity: 25.459

alignment_block:

US-09-303-518D-131 x Q9LWY3

Align seg 1/1 to: Q9LWY3 from: 1 to: 739

78 TACCGAAGTCCGCTTGGCGCAAGAAATATGCGCATGCGGCCCTCGA 127

: :

10 HisArgAlaThrGlyGlyTrpGln..... 17

128 TGAATAATCAAGGAAGGTGAAGCGCTCAAAAAAGGCCAAGTGTGTGAA 177

:
 18 GluGlyGlyGlyArg.....GlnArgProThrAla.....A 28
 178 GACAAAAGAAATCCGGCGGTAGTATTACTGCGCGCTTACAGCAAAAT 227
 :
 28 rArgGluGlyGlyGlyArgArg.....ArgAlaArg 38
 228 CGCGGCTATTCACCGTGGGGAAGGCGGTACTTTCAGTCAGTCGTGATTG 277
 :
 39 ArgGluGlyGlyGlyTrpArgArgPro..... 47
 278 CCCTTGAAGCAACGACGAAATCGAGTTCGAAGCTACGTACCTCTGAACG 327
 :
 48ThrSerArgArgAspAsps 54
 328 CTGCGAAAATTGAGCAGCAAGGTCGCGCAACCTGATTCAATCAGG 377
 :
 54 erGly.....GlnArgArgAlaArgProGlyGlyValThrLeuGln 67
 378 CTTATGAGCTGCGCTTCGACCGGTCCTTCAGCAAAATCCCTCGCGTAG 427
 :
 68 IleLysArgLysProAsnGlyArgValValLeuSerIleProSerTyrTh 84
 428 ATGCCGAGCGCTTCGCCATCTCTCGTCAATGCGATGGACACCAATCCGCTG 477
 :
 84 rThrThrAlaGlnProHis.....HisHisThrProAlaP 96
 478 GTCGCGACCTACGCTCATCAATAAGAGCGCGCA...AGACTTCAA 524
 :
 96 roCysArgLeuValGlyArgHisGlnArgArgArgGlyThrLeuAsp 112
 525 ACGCGCGCTGTTGGTATTGACGCGCT.....GACCGAACGTAAA 565
 :
 113 ArgProProSerSerLeuProProArgAlaIleAsp.SerArgArgL 129
 566 TCCATGTGTAAAGCAGCAGCGCGTCTCGTCTGAAATGCTGCC 615
 :
 129 euHis.....AspHisProSer...AsnProAsp 137
 616 ATATCGAAACACATGAATT.....GGCGCGCGCATCTCTCC 653
 :
 138 AsnIleProLeuProLeuProThrLeuProProHisPro... 153
 654 CGGCTTGAGTCGACGACATTCATTTCATCGAGCGCTCGCGCGAATA 703
 :
 154GlyThrLeuIleArg.....ArgProSerA 162
 704 AAACCGTGTGACCATC.....AATTATCAAGACGTGATTGCT 741
 :
 162 laGluValTrpValValThrSerAlaGlyGluGluGluValIleGly 178
 742 ATCGGACGTTTGTTCGTACAGCGCGCTGAA...TACCGAGCGCGTGG 787
 :
 179 ValGlyGluAspAlaHisGlyAlaArgIleMetLysHisIleGlnArgGl 195
 788 TTGCTTTGGCGCGCTGCAAGTCAACAAACGCGCTCTTGGTACCCTG 837
 :
 195 yCysLeuLysGluAlaAlaSerGlyAspAlaValTrpArgCysAlaA 212
 838 TTGGGTGCGAAGGTCTCAACTTACCCTCGCGCGAATTGGT..... 878
 :
 212 spGly.....ArgArgArgSerGlyAsnGlyAsn 221
 879 ...TGACGCGGAACACCGGTGATTTCGGTTCGGTATTGAA...CGGTG 922
 :
 222 AspGluArgArgGlnCysSerArgArgTyrArgValGluValGluVa 238
 923 CGATTGCACAAGCGCGCATGATTATTGGGACGCTACCAACAATCAGATT 972
 :
 238 lGluCysLeu.....SerLeuProLeuArgGluA 248
 973 TCCGTTATCGAAGA.....AGGCCGACGAAAGAGCTGTTCGG 1010
 :
 :


```
751 AACGTCGATACCAATACGCTTGTGATATGATGTCACACAGGTTTAA 702
      ::::|||||
2104 hrThrSerGluSerValThrSer.....ThrVal.... 2113
701 TTCGCGCCGACTGGCTCGATGAATGAATGTCGTGCCACTCAAGCCGCG 652
2113 ..... 2113
651 AGGATGCGGGCGGCAAAATTCATGTGTTTCGATATGTCGACATTTTCAG 602
      ||| :||| ::::|||||
2114 .....ProGluThrSerLysSerThrValLysSerGluAlaP 2127
601 ACGGACGCTCT...CGCGCTCTCTTTACACACATGATGTTTACGTTGCG 555
      ||||| :||| :||| :|||
2127 roValThrSerThrSerProThrGluValHisThr..... 2138
554 GTCAGGCGGCTCAATACCAACAGCCGCTTTGAAGTCTTCGCGCGCTTC 505
      ::: :||| :||| :|||
2139 .....SerSerGluThrLysProSerLeuSerAlaSer..... 2150
504 TTTGATGATGACCGTAGGTCG.....GCAGCCAGCG 473
      ||| :||| :|||
2151 .....ThrThrGlyAspThrAsnSerThrThrProSerThrSerS 2164
472 GATTGTTGTCATCGATTCAGCAAGATGCGGAACGCG...TCGGCATCT 426
      :||| :||| :||| :|||
2164 erLeuAlaSerValLysSerThrSerAlaProGluGlyThrSerAlaSer 2180
425 ACGGACGAGATTTCGTCGACGAGCGGTCGGAAGCCAGCTCCATAGCC 376
      :||| :||| :||| :|||
2181 ValAlaProValLysLeuSerSer...LeuSerProAspValSerGlnPr 2196
375 TGATTGAATCAGGTTCGGCGCACTTTTCGCTGCTCAATTTGCCAGCG 326
      ||| :||| :||| :|||
2196 oSer.....ThrLysThrPheAsp.....AlaThrG 2205
325 CTTGAGTACGAGTTCGACGCTTTCGCGGCTTTCGCGGCTTTCGCGGAT 226
      ||| :||| :||| :|||
2205 luSerSerThrValGlnAlaSerGluThrSerSerGlyThrSerValLys 2221
275 ATCAAGTACTGACTGAGTACGCTTTCGCGGCTTTCGCGGCTTTCGCGGAT 226
      ||| :||| :||| :|||
2222 SerThrSerGluProGluSerHisValThrLysLeuSerIleThrSerSe 2238
225 TTTGCTGAGCGCGGCGCAGTAACTACTAGCCCGGATTCCTTTTCTCT 176
      ||| :||| :||| :|||
2238 rAsnProSerSerSerValProValThrSerPro.....LysS 2251
175 CAACACGACTTGGCTTTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCATC 126
      ||| :||| :||| :|||
2251 erThrProThrValProGluSerThrGluGlnProThrSerThrPro 2267
125 GAGGGGCGGATCGGACATATTCGCGGCAAGCAAGCGGCTTCG 81
      :||| :||| :||| :|||
2268 SerGlyGlnSerLeuThr.....ProMetAsnSerAsnSer 2279
seq_name: sp_invertebrate.Q23804
```

```
seq_documentation_block:
ID Q23804 PRELIMINARY; PRT; 749 AA.
AC Q23804;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Chironomidae; Chironomidae; Chironominae; Chironomus.
DE SPID PRECURSOR (FRAGMENT).
GN BR2.1.
OS Chironomus tentans (Midge).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Chironomidae; Chironomidae; Chironominae; Chironomus.
OX NCBI_TaxID=7153;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RX MEDLINE=92262483; PubMed=1584794;
RA Wieslander L., Paulsson G.;
RT "Sequence organization of the Balbiani ring 2.1 gene in Chironomus
tentans.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:4578-4582(1992).
DR EMBL; M89909; AAA28264.1; -.
KW Signal.
FT SIGNAL 1 19 POTENTIAL.
FT NON_TER 749 749
SQ SEQUENCE 749 AA; 83564 MW; 606367E2800C45BD CRC64;

alignment_scores:
    Quality: 114.50      Length: 476
    Ratio: 0.500        Gaps: 26
    Percent Similarity: 48.109      Percent Identity: 23.109

alignment_block:
US-09-303-518D-131 x Q23804 ..
Align seg 1/1 to: Q23804 from: 1 to: 749
```

```
12 CAAAAAGGTCTAAATCTCCCATCGCGGCGAGAC.CGGAGCAAGTCATT 60
      :||| :||| :||| :|||
317 GluLysArgSerLysSerGlySerArgProGluLysArgSerLysSerGl 333
61 TATGACGCGCGCCCATACCGAAGTCGCGTTCCTTGCGGAAGAATATGT 110
      :||| :||| :||| :|||
333 ySerArgProGluLysArgSerLysSerGlySerArgProGluLysArgS 350
111 CGGCATCGCCCTCTCATGAAATCAAGGAAGTGAGCCGCAAAAAG 160
      :||| :||| :||| :|||
350 erGluSerLysProArgProGluLysArgSerLysSerGlySerArgPro 366
161 GCCAAGTGTCTTGAAGACAAAAGATCCGGCGGTAGTATTTACTGCG 210
      :||| :||| :||| :|||
367 GluLysCysGlySerLysMetArgVal.....LeuAlaGl 379
211 CGGGCTTCAGCAAAATCGCCCTATTACCGTGGCGGAAAGCGGCTACT 260
      :||| :||| :||| :|||
379 uLysCysAlaAspArgLysGlyArgPheSerAlaSerLysCysArgCysP 396
261 TCAGTCAGTCGTGATTGCGGTTGAAGCAACGACG.....AAA 298
      :||| :||| :||| :|||
396 heSerArg.....ProSerLysSerGlySerArgProGluLysArg 409
299 TCGAGTTCGACGCTACGTACCTGAAGCGTGGCAAAATTCAGCAGCGA. 347
      :||| :||| :||| :|||
410 SerLysSerGlySerArgProGluLysArg.SerLysSerGlySerArgp 426
348 .....AAAAGTCGCGCAACCTGATTCATCAGGCTTATG 383
      :||| :||| :||| :|||
426 roGluLysArgSerGluSerGlyProArgProGlu.....LysArgSer 440
384 GACTCGGCTTCGACCCCTCGCTTCAGCAAAATCCCTCGCTAGATGCGG 433
      :||| :||| :||| :|||
441 LysSerGlySerArgProGluLysCysGlySerLysMetArgValLe 457
434 AGCCGTTCTCGCATCTTCGTCATCGGATGGACACCAATCCGCTGGCGCC 483
      :||| :||| :||| :|||
457 uAlaGluLysCysAlaAspArgLysGlyArgPheSerAlaSerLysCysA 474
484 GACCCCTAC.....GGTCATCATCAAGGAAGCCGCGGA 515
      :||| :||| :||| :|||
474 rgCysPheSerArgProSerLysSerGlySerArgProGluLysArgSer 490
516 AGACTTCAACCGCGGCT.....GTTGGTATTGAGCCGCC 550
      :||| :||| :||| :|||
491 LysSerGlySerArgProGluLysArgSerLysSerGlySerArgProGl 507
551 TGACCGAAGCTAAAT.....CCATGTCTGTAAAGCAGCA 585
      :||| :||| :||| :|||
507 uLysArgSerLysSerGlySerArgProGluGluCysGly...SerAlaM 523
```

```
586 GCGCAGAGCGTCGCTCTGAAATCTGCTGCAATATCGAAACACATGAAATT 635
|||||
523 eArgArgAlaLeuAlaGluLysCysAlaAspArgLysGlyArg..... 537
636 TGGCGCGCGCATCTGCGCGGCTT.....GAGTGGCAGCGACA 673
538 pheSerAlaSerLysCysArgCysPheSerArgProSerTrpSerGlyL 554
674 TTCAATTCATCAGGCGAGTCGCGGGAATAAAACCGTGTGGACCATCAAT 723
554 eLysProGluLysArgSerArgSer..... 562
724 TATCAAGACGCTGATTCCTATCGGACGTTTGTTCGTAACAGCGCTGAA 773
563 .....GlySerArgProGluLys 568
774 TACCGAGCGGTGGTTCCTTGGGGCGCTGCAAGTCAA...CAACCGC 820
569 ArgSerLysSerGly.....SerArgProGluLysArgSerLysSerG1 583
821 GCCTCTGGTACCGTTTGGTGGCGAAGTGTCTCAACTTACCGCGCGC 870
583 ySerArgSerGluLysCysGly.....SerLysMetArgArgV 596
871 GAATTGTTACCGCGCACAAACCGCGTATTCGGTTCGGTATTGAACGG 920
596 alLeuAlaGluLysCysAlaAspArgLysGlyArgPheSerAlaSerLys 612
921 TGGGATTGCAACAGCGCGCATGATTATTGGGACGCTACCAATCACA 970
613 CysArgCysPheSerArgProSerLysSerGlySerArgProGluLysAr 629
971 TTTCCTGATCGAAGAGCGCGCAGCAAGAGCTGTTCGGTGGGTGGCG 1020
629 gSerLysSerGlySerArgProGluLysArgSer..LysSerGlySerAr 645
1021 C...CGCAGCGCAAAATCTCATCAGCGCACCACTCTCGGCCATTT 1067
645 pProGluLysArgSerLysSerGlySerArgProGlu..... 657
1068 CCTAAACAACTCTTCAAGTTCACGA.....CAGCGCTCAACGGCG 1111
658 ....LysArgSerLysSerGlySerArgProGluLysArgSerLysSer 672
1112 GCGACCGCGCATGTTACCGCATCGCACATTATGAGCGCGTAAATGCGGTTG 1161
673 GlySerArgProGluLysArgSerLys..... 681
1162 GACATCTCGCTACCTTGTCTTTGCGCGATTTAATCGTCGCGCATACCGA 1211
682 .SerGlySerArgProGluLysCys.....GlySerAlaMet.... 693
1212 CAGCGCGCGCTTGGGTTGCTTGGNAATGGACGAGAGACC..... 1255
694 .....LysArgThrGluGly 698
1256 TCGCTTTGTCAGCTCTGCTGCGCGCAATACGAATACGCGCGCGCTG 1305
699 GluLysCysAlaArgLysAsnGlyArgPheAsnSerLys.....ArgCy 713
1306 T...TGGCGAAAGTCGTGGAAACA 1327
713 sThrCysThrSerValGlyLysPro 721
seq_name: sp_rvirus:Q98WW2
seq_documentation_block:
ID Q98WW2 PRELIMINARY; PRT: 1504 AA.
AC Q98WW2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
```

```
DE GAG POLYPROTEIN [CONTAINS: CORE PROTEIN(S) P24].
GN GAG.
OS Avian leukosis virus.
OC Viruses: Retroviral viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ENDOGENOUS AVIAN LEUKOSIS VIRUS 1, EV-1;
RX MEDLINE=21165271; PubMed=11264350;
RA Johnson J.A., Henne W.;
RT "Characterization of endogenous avian leukosis viruses in chicken
embryonic fibroblast substrates used in production of measles and
mumps vaccines.";
RL J. Virol. 75:3605-3612(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ENDOGENOUS AVIAN LEUKOSIS VIRUS 1, EV-1;
RA Johnson J.A., Henne W.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO GAG GENE PROTEIN P24 (CORE NUCLEOCAPSID PROTEIN).
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO
KNOWN AS THE RETROPEPSIN FAMILY.
DR EMBL: AY013303; AAK13200.1; -.
DR HSP: P03322; 2RSP.
DR InterPro: IPR001969; Asp_protease.
DR InterPro: IPR001995; Asp_prot_retrov.
DR InterPro: IPR000721; Gag_p24.
DR InterPro: IPR003308; Integrase_zn.
DR InterPro: IPR004028; Retro_M.
DR InterPro: IPR002156; RNaseH.
DR InterPro: IPR001584; Rve.
DR InterPro: IPR000477; Rvtse.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00607; gag_p24; 2.
DR Pfam: PF02022; Integrase_zn; 1.
DR Pfam: PF02813; Retro_M; 1.
DR Pfam: PF00075; rnaaseh; 1.
DR Pfam: PF00665; rve; 1.
DR Pfam: PF00077; rvp; 1.
DR Pfam: PF00078; rvt; 1.
DR Pfam: PF00098; zf_CCHC; 2.
DR SMART: SM00343; Znf_C2HC; 2.
DR PROSITE: PS00141; ASP_PROTEASE; 1.
DR PROSITE: PS00175; ASP_PROT_RETROV; 1.
KW Aspartyl protease; Core protein; Hydrolase; Polyprotein;
KW RNA-directed DNA polymerase; Zinc-finger.
FT CHAIN 1 175 P19 MATRIX PROTEIN.
FT CHAIN 178 239 P10.
FT CHAIN 240 479 P27 CAPSID PROTEIN.
FT CHAIN 489 577 P12 NUCLEOPROTEIN.
FT CHAIN 578 701 P15 PROTEASE.
FT CHAIN 709 1504 TRUNCATED REVERSE TRANSCRIPTASE-BETA
SUBUNIT.
FT CHAIN 709 1280 REVERSE TRANSCRIPTASE-ALPHA SUBUNIT.
FT CHAIN 1281 1504 TRUNCATED INTEGRASE.
SQ SEQUENCE 1504 AA; 163199 MW; 979779BA89620435 CRC64;
alignment_scores:
Quality: 114.50 Length: 528
Ratio: 0.504 Gaps: 31
Percent Similarity: 42.992 Percent Identity: 23.106
alignment_block:
US-09-303-518d-131 x Q98WW2
Align seg 1/1 to: Q98WW2 from: 1 to: 1504
1 ATGATTAAATCAAAAGGTCTAAATCTGCCCATCGCGGCGAGACCGGA 50
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
410 LeuileLysAlaValGluGlySerAspLeuProProSerAlaArgAlaPr 426
51 GCAAGTCATT.....TATGACGCGCGCGCATACCGGAAG 85
```


426 oValilelleAspCysPheArgGlnLysSerGlnProAspIle... 440
86 TCGCGTTCGTTGGGGAAGAAATATGTCGCGCATCGCCCTCATGAAATC 135
441GlnLysLeuIleArgAlaAlaProSerThrLeuThr 452
136 AAGGAAGTGAAGCGGTCAAAAGGAGGCAAGTGTGTTGAAGACAAAA 185
453 ThrProGlyGluIleLys... 459
186 GAATCCGGCGCTAGTATTTATTCGCGCGGCTTCAGGCAAAATCCCGCTA 235
460TyrValLeuAspArg...GlnLysThrAlaProLeu 470
236 TTCACCGTGGGAAAGCGGTACTTCAGTCAGTCGTCGATTCGCGTTG... 283
471 ThrAspGlnGlyIleAlaAlaAlaMetSerSerAlaIleGlnProLeuVa 487
284 AAGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGAAGCGCTGGC 332
487 lMetAlaValVal...AsnArgGluArgAspGlyGlnThrGlySerGlyGly 503
333 AAAATT... 338
504 ArgAlaArgGlyLeuCysTyrThrCysGlySerProGlyHisTyrGlnAl 520
339GAGCAGCGAAAGTGGCGCGCAACCTGATCAATCAGGC 378
520 aGlnCysProLysArgLysSerGlyAsnSerArgGluArgCysGlnL 537
379 TTATGGAGTGGCTCCGACCGCTCCGTCAGCAAAATCCCTCGCTGAGA 428
537 eucysaspGlyMetGlyHisAsnAlaLysGln...CysArgArg 550
429 TGCCGAGCCGTTGCCATCTTCGTCATGCGATGGACACCA... 470
551ArgAspGlyAsnGlnGlyGlnAr 558
471TCCGCTGGCTCCGACCTACGCT... 494
558 gProGlyArgGlyLeuSerSerGlySerTrpProValSerGluGlnProA 575
495CATCATCAAGAGAGCCGCCCAAGA 518
575 laValSerLeuAlaMetThrMetGluHisLysAspArgProLeuValArg 591
519 CTTCAACGCGGCTGTTGGTATGTAGCGCGCTGACCGCAACGTAATCC... 568
592 ValilelleThrAsnThrGlySerHisPro...ValLysGlnArgSerVal 608
569 ATCTGTGTAAAGCA...GCAGGCGCAGACGTG... 597
608 YrIleThrAlaLeuLeuAspSerGlyAlaAspIleThrIleSerGlu 624
598CGTCTGAAATGCTGCCAATATCGAATC 626
625 GluAspTrpProThrAspTrpProValMetGluAlaAlaAsnProGlnI 641
627 ACATGAATTTGGGGCGCCGATCTCCGCGCTTGAGTGGCAGCAGATTC 676
641 eHisGlyIleGlyGlyIleProMetArgLysSer... 653
677 ATTTCATCAGCAGTCGCGCGGCAATAAACCGTGTGGACCATCAATAT 726
653 653
727 CAAGACGTGATTCATCGGAGGTTTGTTCGTAACAGCGCGCTCGAATAC 776
654 ArgAspMetIleValGlyValIleAsnArgAspGlySerLeuGluAr 670
777 CGAGCGCGTGGTT...GCCTTGGGCGGCTGCAAGTCAACAAACCGC 820
..... 820

670 gProLeuLeuPheProAlaValAlaMetValArgArgGluTyrProA 687
821 GCCTCTTGGCTACCGTTCGGTTCGGAAGGTGTCCTCAACTTACCGCGGC 870
687 rg.....LysArgLeuSerAlaGly 693
871 GAATTCGTTGACCGCGCAACACCGCGTATTCGGTTCGTTGATTAACCG 920
694 ProArgAlaProLeuAspLysPheIleGlyArgAlaThrValLeu 710
921 TGGATTCACAGCGCGCATGATTATTGG.....GACCTACACCA 964
710 alAlaLeuHisLeuAlaIleProLeuLysTrpLysProAspHisThrPro 726
965 ATC.....AGATTTCGTTATGAAAGAGCGGC..... 993
727 ValTrpIleAspGlnTrpProLeu...ProGluGlyLysLeuValAlaLeu 742
994AGCAAGAGCTG...TTCGGCTGGTTCGCCGCGCAGCC 1028
743 ThrGlnLeuValGlnLysGlnLeuGlnLeuGlyHisIleGluProSerLe 759
1029 GGACAATACTCC.....ATCACGCGCACCACTCTCGGCCATT 1066
759 uSerCysTrpAsnThrProValPheValIleArgLysAlaSerGlySer 776
1067 TCCTAAAAACAACACTCTCAAGTTCACGACAGCGCTCAACGGCGCGAC 1116
776 Yr.....ArgLeuLeuHisAspLeuArgAlaValAsnAla..... 787
1117 CGCGCCATGTTACCGATCGGCACCTTATGAGCGCGTAATGCGG...TTGGA 1163
788 ...LysLeuValProPheGlyAlaValGlnGlnGlyAlaProValLeuSe 803
1164 CATCTGCTCT.....ACCTTCTTTCGCGGATTTAATCGFCGCGC 1204
803 rAlaLeuProArgGlyTrpProLeuMetValLeuAspLeu..... 816
1205 ATACCGACAGCGCGCGCTTGGGTTGCTTGGATTTGGACGAGAGAC 1254
817LysAspCysPhePheSerIleProLeuAlaGluGlnAsp 829
1255 CTCGCTTTTGCAGCTCTGCTCGCGCGCAATACGAATACGCGCGCT 1304
830 ArgGluAlaPheAlaPheThrLeuProSerValAsnAsnGlnAlaProAl 846
1305 GTTGGCGC.....AAAGTCGTG 1320
846 aArgArgPheGlnTrpLysValLeu 854
seq_name: sp_invertebrate:Q9BKV7
seq_documentation_block:
ID: Q9BKV7 PRELIMINARY; PRT; 1325 AA.
AC Q9BKV7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PFG3.
GN PFG3.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Myler P.J., Sisk E., Cawthra J., Handley F., Vogt C., Robertson L.,
RA McDonagh P., Ivens A., Nguyen D., Munden H., Stuart K., Worthey E.A.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC084329; AAK31375.1;
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003591; LRR_typ.
DR Pfam; PF00560; LRR; 6.

RL Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
 RN [3]
 RC SEQUENCE FROM N.A.
 RP STRAIN=CV. COLUMBIA;
 RA Theologis A.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AC000098; AAB71465.1; -;
 SQ SEQUENCE 402 AA; 39721 MW; D859630FB5C75466 CRC64;

alignment_scores:
 Quality: 112.50 Length: 438
 Ratio: 0.571 Gaps: 18
 Percent Similarity: 44.977 Percent Identity: 22.831

alignment_block:

US-09-303-518d-131/rev x 023054

Align seg 1/1 to: 023054 from: 1 to: 402

```

1331 TCAATGTTTCCAGCACTTTGCGCAACAGCGGCGC.....TATTC 1291
|||||
20 SerThrTyrProAsnThrTyrSerAsnSerAsnProAsnAlaAlaSe 36
1290 GTATTTCGCCGG...CAGACGAAGTCACAAAGCAGAGCTCTTCGCT 1244
|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
36 rSerMetProSerThrAsnThrThrValProGlnThrSerSerS 53
1243 CCAATTCACAGCAACCCAAAGCTGCGCTGTCGGTATCGCGCAGCAT 1194
|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
53 erThrSerThrThrThrAlaThrGluSerSerSerGlyThrThrAla 69
1193 AATTCGCGCAAAAGC...AAGTAGGAGGATGTCACAAAGCGCATTCGCG 1147
|||||
70 GluSerSerSerThrThrLysSerAlaThrMetSerGlySerThrHi 86
1146 CTATAGTCCCGATCGGTACCATCGCGCGCTGCGCGCCGCTTACGCGCTG 1097
|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
86 sThr.....ThrSerAlaThrAlaSerSerThrAla 98
1096 TCGTGAACCTGAAGAGTTTGTGTTTATAGGAATGGCGGAGGTGGCGC 1047
|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
98 erThrSerThrSerSer..... 103
1046 GTGATGAGTATTTGCGGCTCGCGCAACCCAGCGCAACAGCTCTTT 997
|||||
104 .....TyrSerThrSerTyrSerThrSerSerThrLysThrHrth 117
996 GCTGCGGCTTCTTCGATACGGAATCTGATGCTGAGCTGCCCAAT 947
|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
117 rMetThrGlySerThrIleSer..... 124
946 AATCATCGCGCTTGTGCAATGCGACCTTCAATACCGAAGCGGAATC 897
|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
125 .....ThrThrAlaSerAlaAlaProThrSerThr..... 134
896 ACCGGTGTGCGGTACCAATTCGCGCGGTAAGTTGACACACCTT 847
|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
135 .....AlaSerThrSerThrSerTyrSerThrSerThrSerSe 149
846 CGCACCCAAACGGTACGC.....AAGAGCGCGGTTTGTTCGACTT 806
|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
149 rSerThrLysThrThrValThrGlySerThrIleGlyThrThrAla 166
805 GCAGCGCGCCCAAGCAACCGCTCGGTATTCAGA..... 768
|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
166 erAlaAlaProThrSerThrSerThrThrAlaSerSerAlaSer 182
767 .....CGCCCTGTACGAACACGTCGGATAGCAATCACGCTGTG 727
|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
183 SerThrThrAsnProSerSerGlySerLysProThrAlaMetThrGly... 198
726 ATAATTGATGTCACACAGGTTTATTCGCGCGGACTGGCTCGATGAAT 677

```

```

199 .....ThrThrAlaAsnThrSerProSerAlaProThrSerSer..... 211
676 GAATGTGCGTGCACACTCAAGCGCGGAGGATGCGGCGCCAAATTCATGT 627
212 .....ProSerThrThrAsn 216
626 GTTTCGATATTCAGCATTTTCAGACGCGAGCTGTCGCGCTGCTGCTTT 577
|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
217 SerSerSerThrAlaAlaTyrThrSerSerGlySerLysProThr..... 231
576 ACACACATGATTTTACGTTTCGTCAGCGGCTCAATACCAACAGCCGCG 527
232 .....ThrValThrArgThrThrAlaAsn..... 239
526 GTTTGAAGTCTTGGCGGCTTCTTTGATGATGACCGTAGGTCGCGCAGC 477
|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
240 .....ThrSerSerAlaSerThr.....SerSerAla 249
476 AGCGGATTCGTCATCGCATTCAGGAAGATGGCAACGCGTCGCGCATC 427
|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
250 Ser.....ProThrAsnSerSerThrSe 257
426 TAGCGCAGGATTTTCTGCTGACGCGAGCGGTGCGAAGCGCAGTCCATAAGC 377
|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
257 rThrPro.....ThrAsnSerSerAlaGlySerLysP 268
376 CTGATTGAATCAGGTGCGCGCAGCTTTTTCGCTGCTCAATTTGCCAGC 327
|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
268 ro.....ThrThr 270
326 GCTTCAGGTACGTAGCGTTCGAACTCGATTTCGTCGTCGTCGTCGTCG 277
|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
271 MetThrGlyThrThrThrAsnThrSerSerThrThrThrThrSerSerAl 287
276 AATCAGCACTGACTGAAGTACGCGCTTTTTCGCGCAGCTGAATAGCGGCA 227
|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
287 aSerThrThrLysSerSerSer..... 294
226 TTTTGCCTGAACCGCGCGAGTAATACTACGCGCGGATCTTTTGTCT 177
|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
295 .....SerSerAlaThrAsnSerSerSerGlySerLysProSer 307
176 TCAACACAGCACTTGGCTTTTTCGCGCTTCACCTTCCTTCGATTTCAT 127
|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
308 ThrLeuSerThrThrAlaTyrThrAlaThrThrSerSerProThrAl 324
126 CGAGGCGCGCATGCGCACATATTCTTCGCCAAGCAACGCGACTTCGGTAA 77
|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
324 aGluProSerThrThrThrAlaSerLysProAlaThrSerSerThrProp 341
76 TGGCGCGCGCGTCA 63
|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
341 roAlaProThr 345
seq_name: sp_bacteria:006470
seq_documentation_block:
ID 006470 PRELIMINARY; PRT; 405 AA.
AC 006470;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CEOA.
GN CEOA.
OS Burkholderia cepacia (Pseudomonas cepacia).
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Burkholderia.
OX NCBI_TaxID=292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K61-3;
RA Charron R.L.; Barry J.J.; Burns J.L.;

```

103 GAATATGTCGCATCGCCGCC.....TCGATGA 131
||| ||| :::::|||||
71 GluIysValAspValArgProGlnValSerGlyThrIleValAlaValAs 87

132 RATCAAGAAGTCAACCGCTCAAAAAGGCCAAGTGCTGTTTGAAGACA 181
::::|||:::||| |||||||||:::|||||
87 nPheLysaspGlyAlaLeuValLysGlyAspValLeuPheValIleA 104
::::|||:::||| :|||::|||
182 AAAAGAATPCGGGGTAGTATTACTTCGCGCGCTTCAGCGCAAATCGCC 231
:::|||:::||| :|||::|||
104 sPProArgProTy rGlnAlacl uValAspArgAlaAlaGlnLeuAla 120

232 GCTATTTCAC.....CTGGCGCAAA 251
||| ||| :::::|||||
121 AlaAlaGlnAlaLaarAsnGlyTy rAlaGlnThrAspTrpGlnArGaGl 137

252 GGCGCTACTTCAGTCAGTCGATTGCGGTGAAGCAACGACGAAATCG 301
:|||:::|||:::||| :|||::|||
137 nArgLeuleGlyAspAsnAlaIleAlaLysArgSpTy rAspGlu.Lys 153

302 AGTTCGAACGCTAGCTACCTGAACGCTGGCAAAATTCAGCAGCAANA 351
:::|||:::||| :|||::|||
154GlnAsnAlaAla..ArgGluA 160

352 GTGCGCGCCCAACCTGATTCAATCAGGCTTATGAGCTGCGTTCGCAACC 401
:::|||:::|||
160 lAlyArgGluPr oGluGly..... 166

402 TCCGTTCAAGAAAAATCCCTGCCGTAGATGCGAGCCGTTGCG.....CA 445
||| ||| ::||| |||||
167ArgArgSerArgAlaGlyAsnGlyAlaH 176

446 TCCTTGCTCAATGCGATGACACCAATCCGCTGGCTGCCGACCC.....? 488
| ::::|||:::||| |||
176 sGlnSerArgLeuTy rAlaHisArgAlaGlyValGlyProArgValA 193

489 ..TAGCGTTCATCATCAAAAGCCGCGAGACTTCAAACGCGCGCTGTT 536
|||:::||| |||||::|||
193 laArgGlyAsnHisalaargGlnArg..... ValGlyArgArgVal 207

537 GGTATTGACCGCGCTGAC.....CGACCGTA 562
||| |||
208 GlyArgAlaalaAspAlaGlyIleGlyValAlaAspLeuArgVal... 223

563 AATTCATGTGTATAAGCACGCGCAGACGTGCGCTCTGAAATGCT 612
:::|||:::||| :|||::|||
224ValArgArgArgAlaAspLeuProAlaI 234
||| ||| |||||

613 GCCAATATCAAAACATGAATTGGCGGCCGATCCTCGCGCTTGAG 662
:::|||:::||| :|||::|||
234 leHisGlnArgargAla.....GlnArgProGlnGlyAla..... 245

663 TGCGACGCACATTCATTATCGAGCCAGTCGCGCGAATAAACCGTGT 712
||| ||| |||||
246GlyVArgAlaArgProArgGluArgAsnArgLe 256

DR EMBL; U60315; AAC55161.1; -.

```

seq_name: sp_virus:Q98201

seq_documentation_block:
ID      Q98201      PRELIMINARY;      PRT;      579 AA.
AC      Q98201;
DT      DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT      DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT      DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DT      DT      MC033L.
DE      Molluscum contagiosum virus subtype 1 (MCV1).
GN      Molluscum contagiosum virus, no RNA stage; Poxviridae; Chordopoxvirinae;
OS      Viruses; dsDNA viruses, no RNA stage; Poxviridae;
OC      Molluscipoxvirus.
OC      Molluscipoxvirus.
NCBI_TaxID=10280;
OX      [1]
RN      SEQUENCE FROM N.A.
RP      MEDLINE=96325459; PubMed=8670425;
RX      Senkevich T.G., Bugert J.J., Sisler J.R., Koonin E.V., Darai G.,
RA      Moss B.:
RT      "Genome sequence of a human tumorigenic poxvirus: Prediction of
RT      specific host response-evasion genes.";
RL      Science 273:813-816(1996).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      Senkevich T.G., Bugert J.J., Sisler J.R., Koonin E.V., Darai G.,
RA      Moss B.:
RL      Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DR      EMBL; U60315; AAC55161.1; -.
DR      InterPro: IPR003600; Ig-like.
DR

```

DR SMART: SM00410; IG-like; 1.
SQ SEQUENCE 579 AA; 62626 MW; A37930DF92D311A3 CRC64;

alignment_scores:
Quality: 112.00 Length: 473
Ratio: 0.563 Gaps: 23
Percent Similarity: 42.072 Percent Identity: 21.142

alignment_block:
US-09-303-518D-131 x Q98201 ..
Align seg 1/1 to: Q98201 from: 1 to: 579

115 ATGCGCCCTCGATCAAAATCAAGGATGAGCGCTCAAAAAA..... 159
1 MetArgProHisValLeuThrLeuAlaThrCysAlaLeuArgAlaLe 17
160 .GGCAAGTCTGTTGAAGACAAAAGATCCGGGGCTAGTATTTACTG 208
17 uAlaGlnValIleAspGluHisGluHisSerGluProProValSerThr 34
209 CGCGGCTTCAGGCAAAATCCGCTATTACCGTGGCGGCAAAACCGGTA 258
34 rpProAspMetSerTyrIleValAlaGluMetArgSerAspValLeu 50
259 CTTCAGTCAGTCGTGATTGCGGTTGAAGCAACGACGAAATCGAG..... 303
51 MetArgGlyMet.....LeuAspGlyHisGluHisValArgCysTh 64
304TTGACCGCTAGGTACCTGACGCGGTGG 331
64 rCysValProLysCysAlaCysLeuGluProThrLeuProArgAlaAla 81
332 CAAATATGACGAGCAAGAAAGTG..... 354
81 euGluGlnAlaArgSerArgValLeuAspAlaHisAlaGlyArgValPro 97
355CGCGCAACCTGATTCAATC 374
98 GlyLeuArgAlaProSerArgAlaAlaHisArgArgValValLeuThrAl 114
375 AGGCTTA...TGACTCGCTTCGCACCCCTCGCTTCAGCAAAATCCCTG 421
114 aGlyCysArgPheThrGlnGlyPheProGluProPheGluClyLeuTrp 131
422 CGGTAGATCGCGCGCTTCGCC..... 444
131 alAlaSerAlaGluProGlyAlaHisAspGluPheCysValGlyGlu 147
445ATCTCGTCAATCGGATGGACACCAA 470
148 HisCysAspAlaHisLeuSerSerIlePheCysHisAlaSerThrMe 164
471 TCGCGTGGCTGCGGACCT.....ACGGTCATCATCA 502
164 tProLeuAlaArgAlaProHisSerProProMetValThrPheSerAlaL 181
503 AAGACCGCGGAGAACTTCAACAGCGGCTGTGTATTGAGCGCGCTG 552
181 euSerAlaGlyGluAsnTyrLeuArg..... 189
553 ACCGAACGTAATCCATCGTGTAAAGCAGCAGCGGCA..... 591
190LeuValCysArgAlaSerGlyAlaTyrProProVa 201
592GACGTGCGCT 601
201 lAspThrLeuThrLeuValSerGlnProGlnProGluAspAlaProC 218
602 CTGAAATGCTGCATATCGAAACACATGAATTGGCGGC..... 642
218 ysGluThrTyrAlaGlyThrAsnAlaAspSerThrGlyHisValGlyMet 234

643CCGCA 647
235 AlaCysValArgSerAspAlaLeuAlaGlyAlaAlaCysAlaValGlnHi 151
648 TCCTGCGCGCTTGAGTGGCAGCAGCATTCATTTCATC..... 684
251 sArgGlyValThrThrSerAlaArgIleValLeuValProAlaAsnAspG 268
685 ..GAGCCAGTCGCGCGCAATAAACCCGTGTGGACCATCAATTATCAAGAC 732
268 lAlaLysValGlyAlaTyrAlaAspValAspAlaAspPheTyrAlaAsp 284
733 GTG.....ATTGCTATCGGAGCTTT 752
285 ValProProLeuProGluProGluSerAspSerLeuAlaValHisAlaLe 301
753 GTTCGTAACAGCGCTCTG.....AATACCGAGCGGTGG 787
301 uPheValAlaGlyAsnThrGluLeuTyrValHisGlyThrAlaAlaGly 317
788 TTCCTTGGCGCGCTGCAAG...TCACAAACCGCGCTCTTGGGTACC 834
318 ValProSerAlaSerCysArgCysAspThrArgArgCysThrCysVal... 333
835 GTTTTGGTGGCGAAGGTCTCAACTTACCGCGCGGGAATTGG..... 877
334LeuAlaProAlaThrTrpThrAlaG 342
878TTGACGCGGACAAACCGCTGATTTCGGTTCGGTAT 913
342 lYValValArgGluLeuAlaAlaAlaAla..... 352
914 TGAACGGTGGATTGCACAAAGCGCGCATGATTATTGGAGCGTACCAC 963
353HisAspLeuLeuAlaValLeuAspValHisA 364
964 AATCAGATTTCGTTATCGAAGAAGCGCGCAGCAAGAGCTTTCGGCTG 1013
364 lAspGlyLeuAlaLeuAsnArgSerSerMetGln.....ValTyrAla 378
1014 GGTGGCGCGCAGCGCGCAATAACTTCCATCAGCGCGCACCTCTCGGCC 1063
379 GluCysGlyProAlaGly..... 384
1064 ATTTCTAAACAACTCTTCAAGTTTCAGACAGCGCTCAACGGCGC 1113
385ArgA-gLeuArgValHisAsnThrGlyThrArgArg. 396
1114 GACCGCGCATGGTACCGATCGCCTTATGAGCGCGTAATGCGGTTGGA 1163
397GlnArgValCysAlaArgGlyAlaCys... 405
1164 CATCTGCTACCTT 1178
406 GluProAlaTyrLeu 410
seq_name: sp_bacteriap:Q9RYP2
seq_documentation_block:
ID Q9RYP2 PRELIMINARY; PRT; 376 AA.
AC Q9RYP2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ADENINE DEAMINASE-RELATED PROTEIN.
GN DRA0268.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;

```

RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamatheva J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RL radiodurans R1."
RL EMBL; AE001863; AAF12376.1; -.
DR TIGR; DRA0268; -.
KW Complete proteome.
SQ SEQUENCE 376 AA; 39845 MW; 7AB7FF32F8C45651 CRC64;

alignment_scores:
  Quality: 111.50      Length: 290
  Ratio: 0.961        Gaps: 15
Percent Similarity: 40.000 Percent Identity: 26.897

alignment_block:
US-09-303-518D-131 x Q9RYP2
Align seg 1/1 to: Q9RYP2 from: 1 to: 376

639 CGCGCGCGCA.....TCTCGCGCTTGAGTGGCAGCCACA 673
|||||
35 ArgProAlaLeuArgGlyAspSerAlaArgLeuAlaGlyHisAlaH 51
674 TTCATTTCAT.....CGAGCCAGTCGGCGGAATAAAACCGTGGACC 717
|||||
51 SHisThrHisPheGlyAlaAlaGlyArgThrGly.....62
718 ATCAATTATCAAGACGTGATTGTCATCGGACGTTTCTGTAACAGCGCG 767
|||||
63 .....ThrGlyArgPro 66
768 TCTGATATACGAGCGGTGTCCTTGGCGGCTGCAAGTCAACAAC 817
|||||
67 AlaAlaGlyProAlaGlyGlyGlnSerGlyAlaGlyGluProArgArgVa 83
818 CGCG...CCCTTCGGTAC.....CGTT 837
|||||
83 lAlaHisProAlaLeuTyrSerGlyGlyArgAlaAspArgAlaGlyArg 100
838 TTGGTGCGAAGGTGTCACACTACCGCGCGAATTGTTGACGCGGA 887
|||||
100 laGlySerAlaAlaGlyAlaArgTyrArgGluIleLeuGly...ProAla 115
888 CAACCGCGTATTCGGTTCGGTATTTGACGGTTCGATTCACAGGCG 937
|||||
116 ArgProArgGlyAspArgProTyrValSerArg.....ArgLeuArgAl 130
938 CGCATGATTATTGGGACGCTA.....959
|||||
130 aAlaAlaArgLeuAlaThrLeuArgGlyAlaGlyAsnLeuArgGlyArg 147
960 .....CCACAATCAGATTTCGTTAT 980
|||||
147 rgGlySerAlaAlaArgArgAspAlaProProAla.....ArgTyr 161
981 CGAAGAGCGCGAGCAAGAGCTGTTGCTCGCTGGT.....1016
|||||
162 ArgArgArgProArg.....AlaArgLeuGlyArgGlyHisLeuAr 175
1016 .....1016
175 gSerAlaArgAlaLeuAlaHisAlaProAspValSerArgProAspArgH 192
1017 .....TCGCGCGAGCGGCAAAATATCCATACGCGCACCACTTCGG 1061
|||||
```

Align seg 1/1 to: Q9N5K0 from: 1 to: 1203

```

1343 CAGCCCTTCCTTCAATGTTTCAGACACTTTCGCGCACACCGCGCGTA 1294
1344 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
411 GlnGlySerProSerThrValSerGlnThr.....ThrProGlyValSe 425
1293 TTCGTATTTCGCGCGGAGAGAGAGTGCACAAAGCGAGGTCTCTTCGT 1244
425 rSerAlaSerThrGlyMetThrThrSerGlnAlaSerLeuArgSerThrG 442
1243 CCAATTCC.....AGCAACCCAAAGCGTGCCTGCGGTATCCGCCG 1200
1244 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
442 InAsnSerValThrSerThrProGlySerPheThrValThrSerThrPro 458
1199 ACGATTAAATCGCCAAAGCAAGGTAGGAGGATGCCAGGATCCACGCAATAC 1150
459 ThrIleThrSerThrGln.....GlyAlaMetAlaSerSerSe 472
1149 GCGCTCATAGTCCGATCGCATACCATGCGCGGTGCGCGCGGTGACGG 1100
472 rAsnSer.....ProThrSerThrGlnAla..... 480
1099 CTGTCGTGAAGTGAAGAGTTGTTTATAGGAATCGCGAGAGTGGT 1050
480 ..... 480
1049 CGCGTATGAGTATTTTCGCGGTGCGCGCAACCCAGCGAAGACAGCTC 1000
481 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
999 TTTCTCGCGGCTTCCTCGATACGGAATCTGATTGTTGTTGCGTCCCA 950
492 rThrIleAlaProThrSerSerGlnGluSer..... 502
949 AATAATCATCGCGCTGTGTGCAATCGCACCTCAATACCGAACCGGAA 900
503 .....SerSerIleAlaSerSerAlaProSerGlnSerThrThrVal 517
899 ATCAGCGGGTTCGCGGTCAACCAATTCGCGCGGTGAAGTTGAGACAC 850
518 IleThrValSerThrThrAlaThrValSerSerGlyGlnSerThrSerTh 534
849 C.....TTGCGACCCCAAAACCGGTAC 830
534 rPheGlyThrIleGlyThrGlySerSerPheGlySerThrIleG 551
829 GCAAGAGCGCGGTTCGTTGACTGTCAGCGCGCGCCCAAGGCAACCGCGC 780
551 InThrThrGlnGlyThrSerSerPheAlaProIleProSerThrAlaGly 567
779 TCGGTATTCAGCGCGCTGTACGAACAACGTCGATAGCAATACAGTC 730
568 Ser.....SerSerGlnThrProGlySerMetSerSe 578
729 TTGATAATTGATGTCACACAGGTTTATTCGCGCGGACTGCTCATCA 680
578 r.....ThrGlyThrThrValGlyGlnMets 587
679 AATGAATGCGTGCCTCAACGCGCGAGGATGCGGG..... 642
587 exSerSerPheGlnProThrAlaProThrSerLeuGlyThrIleMetThr 603
641 ....CCGCCAAATCATGTGTT...TCGATATTGGCAGCAATTTTCAGACGG 598
604 ThrProGlyThrSerSerIleProSerIleSerThrSerValAsnSerG 620
597 CAGCTCGCGCGCTGCTGCTTACACACATGGATTTTACGTTCCGGTCAGC 548
620 ySerSerThrIleGlySerThrValThrGlnAlaProSerSerThrS 637
547 GCGTCAAT.....ACCAAC 534
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

637 erMetGlyProSerProSerGlnSerThrAlaGlySerThrMetThrSer 653
533 AGCCGCGTTCGAAGTTCGCGCGCTTCTTTGATGATGATGATGATGATGATG 484
654 AlaProProValThrSerSerSerAla.....AsnThrGlySe 667
483 GGCAGCAGCGGGA...TTGGTGTCCATCGCATTCAGCAAGATGGCG... 441
667 rThrSerSerGlyThrThrValSerValGlnThrThrGlnValSerSerT 684
441 ..... 441
684 hrThrSerProValAlaSerSerSerSerGlnMetThrSerThrGlnGln 700
440 ...AACGCTTCGCTCTACGCGAGGATTTTCTGCTGAACGCGAGGTGGC 394
701 ProSerGlySerSerSerSerIleGlySerThrValAsnGlnGlySerSe 717
393 AAGCGCAGTCCATAAGCCTGATTGATCAGGTTCGCGCGCACCTTTTCGC 344
717 rSerValThrThrGlnProPro..... 724
343 TGCTCAATTTTCCAGCGCTTCAGGTACGTTCGAGTTCGAACTCGATTTCG 294
725 .....AlaSerSerArgSerThrAlaSerGlnGlySerSerSer 737
293 TCGTTGCTTCACACGCAATCAGCTGACTGAACTGACG.....CGCTT 250
738 AlaGlnProIleAlaSerSerSerThrMetGlySerThrAlaGlySerSe 754
249 TTCGCGCAGGTGAATAGCG..... 231
754 rSerProGlnProThrAlaSerSerThrSerValProSerSerThrGlyA 771
231 ..... 231
771 laThrSerSerGlySerThrValGlySerSerThrMetGlySerThrGln 787
230 GCGATTTTCCTGAGCGCGCGCAGTAACTACTACGCGCGGATTCCTTTT 181
788 SerSerLeuProSerSerThrMetThrAsnThrGly..... 799
180 GTCTTCAACAGCACTTGGCGCTTTTTCAGCGCTTCACCTTCCTTGATTT 131
800 ..SerThrGlySerThrValThrAsnGlnLeuAlaSerSerSerThrThrYrG 816
130 TCATCGAGGCGCGCATCGGACATATTCCTCGCAAGCAACGCGACTTCG 81
816 lyAlaSerSerThrGluProIleAlaSerSerThrAlaAsnProGlySer 832
80 GTAATGGCGCGCGCTCATATAATGACTTGTCTCCGCT 45
833 SerThrSerGlyGlnThrAlaValThrThrGlnGly 844

```

seq_name: sp_bacteria:P94909

seq_documentation_block:

ID P94909 PRELIMINARY; PRT; 529 AA.

AC P94909;

DT 01-MAY-1997 (TrEMBLrel. 03, Created)

DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)

DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)

DE HYPOTHEtical 58.7 KDA PROTEIN (FRAGMENT).

OS Microbacterium ammoniaphilum.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Micrococcales; Microbacteriaceae; Microbacterium.

OX NCBI_taxid=42460;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 15354;

RX MEDLINE=96257250; PubMed=8654988;

RA Striebel H.M., Seiber S., Jarsch M., Kessler C.;

RT "Cloning and characterization of the Mami restriction-modification

64	GACGCCCGGCAT	TACCGAAGTCGC	GTCTTGGCGAA...	GAATATGT	110
					...
159	AspGlyLeu	GlulIleAsp	GlusLeu	ThrGlu	AlaAspPro
111	CGGCATCGCC	CCGCGATGAA	AATCAAGGAGGT	GAGCCGTCA	AAAGAAG
175	lValLys	GlnProGly	AspGlnValMet	SerGlySerPhe	ValValAlaG
161	GCCAGTCGCT	GTGTTGACAC	AAAAAAGAAAT	CGCGCTAGT	ATTACTCGC
192	lyGlyGlyAla	Phe..Alga	ThrLysValGly	ArgGluAla	TyrAlaA
211	CC.....
208	laGlnLeu	AlaGluGlu	AlaSerArgPhe	ThrLeuVal	GlnSerGluLeu
213	GGCTTCAGG	CAAAATCC	CGCTATTCA	CGGTGGCGAAAGCGCG
225	ArgSerGly	IleSerThr	IleLeuLys	TyrValThr	TrpMetValPr
257	TACTTCAGTC	AGTCGTGAT	TCGCTTGAG	CAAGCAGC
241	oThrAla	IleGlyLeu	IleLeuSerGln	LeuLeuAla	TyrAspAsnSP
295	...CAATCGAG	TTCTGACGC	TACGTACC
258	LeuAspAsp	SerIleAla	ArgThrVal	GlyIleVal	ProMetValPr
321	TGAAGCGCT	GGCAAAAT	CAGCAGCG	AAAAAAGT	CGCGCAACCTGATTC
274	oGluGly	LeuValLeu	LeuThrSer	ValAlaPhe	AlaIleGlyValVala
371	AATCAGCCT	TATGACTGC	CTTCGTCG	CACCGCT	CGGTTCAGCAAAATCCCT
291	rgLeuGly	ArgLysGln	CysLeu.....	ValGlnGlu	LeuPro
303


```

427 CTACGGCAGGAGTTTGTCTGAACGACGGGTGCGAAGCGCAGTCCATAAG 378
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
821 erSerSerValSerThrThrThrThrThrThrThrThrThrThrThrThr
377 CCTGATTCAATCAGTTCGGCGCGCAGCTTTTCGCTGCTCAATTTGGCCAG 328
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
838 Pro.....PheLeuPheGlyAlaProGlnAlaSerAl 848
327 CGCTTCAGGTACGTAGCTTCGAACCTGATTCGCTGCTG.....CCTT 284
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
848 aAlaSerPheThrProAlaMetGlySerIlePheGlnPheGlyLysProp 865
283 CAACGGCAATCAGCAGTCACTGAAGTAGCGCTTTTTCGCCACGCTGAATA 234
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
865 roAlaLeuProThrThrThrThrThrThrThrThrThrThrThrThr
233 CGCGCGATTTTCCTGTAAGCGCGCGCAGTAAATACTAG.....V. 195
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
880 .....LeuProThrAlaValProThrAlaThrSerSerAlaAl 893
194 .....CCGGATCTCTTTTGTCTTCAACAGCAGCTTGGCTTTTGA 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
893 aAspPheSerGlyPheGlySerThrLeuAlaThrSerAlaProAlaThrS 910
151 CGGCTTCACCTTCCTGATTTCATCGAGGGCGCATGCCGACATATCT 102
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
910 erSerGlnProThrLeuThrPheSerAsnThrSerThrProThrPheAsn 926
101 TCGCAGCAACGACGACTTCGGTAATCGCGCGCGCTCATAAATGACTTG 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
927 IleProPheGlySerSerAlaLysSerProLeuProSerTyrProGlyAl 943
51 CTCGGGTCTCGCGCGCATGGGC 30
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
943 aAsnProGlnProAlaPheGly 950
seq_name: sp_invertebrate:P91365

```

```

seq_documentation_block:
ID P91365 PRELIMINARY; PRT; 2232 AA.
AC P91365;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE K06A9.1 PROTEIN.
GN K06A9.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP
RC STRAIN-BRISTOL N2;
RA Geisel C., Gattung S.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS (A AND B) MAY BE PRODUCED BY
CC ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE SHOWN IS THAT
CC OF THE A FORM.
DR EMBL; U08046; AAC70889.1; -;
DR EMBL; U08046; AAC70890.1; -;
KW Alternative splicing.
FT VARSPLIC 842 866
FT FT
FT VARSPLIC 870 1051
FT FT
FT PYPQSTSPVESSTTPSPGSGFTLTSTSPSPSQSTTIGST
FT QGSTSPGLTSTSEMTSGGSTQTQPGSTGTVTPSTVSDST
FT SSGSTVTVGSTEGSSPIPTSTQNTNPSTSGSSMSQTPO
FT SQGSTSPVESSTGATSSGSPCTLSISPSPSSTIGS
FT SQGSTSPWSTISGSTE -> KEIDQTAINTYKTYFNAL
FT LVASKLNESLITGIDNFYAGLNDHQYPTDDYNGIKS
FT VPFDIGTDDIDLDLKDVKSLATADWTPPADQTCMFI
FT SAAPEDYGGTITKYTFYFVTVGVVGAKSIPGLSIDK
FT NIVITNTMNDRADSAVVSKLLELPTA (IN ISOFORM

```

```

FT SQ SEQUENCE 2232 AA; 213840 MW; 08D69FA638E14CC8 CRC64;
B).
alignment_scores:
Quality: 108.50 Length: 453
Ratio: 0.522 Gaps: 16
Percent Similarity: 45.916 Percent Identity: 20.530
alignment_block:
US-09-303-518D-131/rev x P91365 ..
Align seg 1/1 to: P91365 from: 1 to: 2232
1340 CTTTCCTTCTCAATGTTTCCAGCACTTTGCGCAACAGCGGCCG..... 1296
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
502 ProSerSerProGlyThrMetSerThrVal.....SerGlyProThrGln 516
1295 .....TATTGCTATTTCGCCGGCGCAGAGAGTGCACAAAGCGAGGT 1253
516 ySerThrValThrValProGlySerSerThrSerProAlaProSerS 533
1252 CTTTCCTTCTCAATGTTTCCAGCACTTTGCGCAACAGCGGCCG..... 1203
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
533 erSerProAsnProSerSerProAlaSerThrGlySerThrIle... 548
1202 CCGCAGATTAAATCGCGGCAAAAGAGGTAAGCAGGATGTCACACGGCAT 1153
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
549 ...ThrIleSerGlySerSerIleValSerThrValSerGlySe 564
1152 TAGCGCTCATATAAGTCCGATCGGTACCATGCGCGGTGCGCGCGCTTGA 1103
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
564 rThrValSer.....GlySerThrGlyThrSerGlnSerThrL 577
1102 CGGCTGCTGTAAGTGAAGACTTTTGTAGGAAATGGCCGAGAGTG 1053
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
577 euAlaSerSerThrAla..... 582
1052 GTGCGCGTCAATGAGTATTGTCGGGTGCGCGCAACCCAGCCGCAACAG 1003
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
583 .....ThrProGlySerSerSerThrValProSerSe 593
1002 CTCCTTGTGCGGCTCTCTTCGATAACGAAATCTGATTGTGGTAGCGTC 953
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
593 rSerSerProGlnProSerSerGlnSerPro..... 603
952 CCAATAATATCATGCGCGCTTGTGCAATCGCACCGCTTCAATACCGAACCG 903
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
604 .....AlaProAsnThrGlySerThr 610
902 GAAATCACGCGGTGTCGCGGTCAACCAATTCGCGCGGTGAAGTTGAGA 853
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
611 ThrProSerGlnThrSerSerGlnSerProSerProSerMetAsnProSe 627
852 CACCTTCGCAACCC.....AAACGGTAGCAGAGAGCGCGC 818
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
627 rSerSerThrProThrGlySerSerGlnSerThrIleThrProGluGlyS 644
817 GTTGTGTTGACTGACGGCGCGCCCAAGGCAACCGCGCTCGGTATTC... 771
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
644 er.....ThrAlaSerSerProThrGlySerThrGlySerThrPheSer 658
770 ...AGACGGCTGTGTACGAACAAACGTCGATAGCAATCAGCTCTTGATA 724
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
659 ValAlaThrGluValThrSerGlnSerThrVal..... 669
723 ATTGATGTCACACGGGTTTTATTTCGCGCGCACTGGCTCGATGAATGAA 674
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
670 .....ProSerGlySer..... 673
673 TGTGCTGCCACTCAAGCGCGCAGGATCGCGGCG.....CCAATTCATGT 627
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
674 ..SerLeuGlyThrGlnSerThrAsnSerSerProSerProSerLeu 689

```

DR EMBL; AF195122; AAF24255.1; -.
DR HSP; P18314; 1FWF.
DR InterPro; IPR001924; Urease.
DR Pfam; PF00449; urease; 2.
DR Pfam; PF02802; urease_C; 2.
DR ProDom; PD002467; Urease; 2.
DR PROSITE; PS01120; UREASE_1; 1.
SQ SEQUENCE 569 AA; 61425 MW; F543AG79D3BBF1DA CRC64;

alignment_scores:
Quality: 108.00 Length: 373
Ratio: 0.643 Gaps: 20
Percent Similarity: 45.040 Percent Identity: 22.788

alignment_block:
US-09-303-518D-131 x Q9RFF2 ..
Align seg 1/1 to: Q9RFF2 from: 1 to: 569

121 CCCTCGATGAAATCAAGGAAGGTGAAGCCGTCAAAAAGGCCAAGTGCT 170
||||| : : : : : ||||| : : : : :
116 ProGlyThrGluIleAlaGlyGlu.....GlyArgIleLe 128
171 G.....T 172
128 uThrAlaGlyGlyMetAspAlaHisIleHisPheIleCysProGlnGln 145
173 TTGAAGACAAAAGAAATCCGGCGGTAGTATTT.....ACT 207
:
145 leGluAspSerLeuHisSerGlyIleThrMetLeuGlyGlyThr 161
208 GCGCGGCTTCAGCAAAATCGCGCT.....AT 236
: : : : : ||| : : : : : : : : : : : : : : :
162 GlyProAlaHisGlyThrLeuAlaThrThrLeuHisAlaGlyAlaLeu 178
237 TCACCGT.....GCGAAAAGCGGTACTTCAGTCAGTCGTGA 274
||||| : : : : : ||| : : : : : : : : : : : : : : :
178 aHisAlaAspAlaAlaGlyGlyArgLeuProAspGlnProArgL 195
275 TTGCGGTTGAAGCAAGCACGAAATCGAGTTGAA..... 309
:
195 euArgGlyGlnGlyCysGlnProAlaArgAlaArgGlyAlaGly 211
310CCTACGTACCTGAACCGCTGCCAAATTCAGCAGCAAAA 350
||||| : : : : : ||||| : : : : : : : : : : : : : : :
212 ProArgGlyArgLeuValProGluAlaAlaArgGlyLeuHisAl 228
351 AGTCCGCGCAACCTGATTCAATCAGCGCT.....ATGAGCTCGCTT 393
:
228 aCysGlyHisArgLeuLeuProValGlyArgArgProHisGlyCysAlaG 245
394 CG.....CACCGCTCCGTTTCAGCAAAATCCCTGCGTAGATGCCGA 434
: : : : : ||| : : : : : : : : : : : : : : :
245 lyAspAspProHisArgHisAlaGluArgGluArgLeuArgGluHis 261
435 GCCGTTCCCATCTTCGT.....CAATCGCATGGACACCAATCGC 475
||| ||||| ||| ||| ||||| : : : : : : : : : : :
262 AlaGlyArgHisArgAlaHisAspProCysValProHisArgGlyC 278
476 TGCTGCGCGACCTACGCTCATCAAGAACGCGCGAAGACTTCAAA 525
: : : : : ||||| : : : : : : : : : : : : : : :
278 sGlyArgArgProCysAlaArgHisHisGlnGlyGlyArgGlu 295
526 CCGGCGCTGTT..... 536
: : : : : ||||| : : : : : : : : : : : : : : :
295 rgHisProValLeuAspGlnSerHisHisAlaLeuHisArgGlnTyrGly 311
537GGTATTGAGCGCGCTGACCGCAACGTA 562
312 GlyGlyAlaSerArgHisAlaAspGlyLeu...ProProSerArgProLe 327
563 AAATCCATGTGTGTAAGCAGCAGCGCGCAGCTGCGCTCTGAAATGCT 612

626 GTTTCGATATTGGCAGCATTTTCAGACGGCAGGTCTGGCGCTGTGCTT 577
: : : : : ||||| : : : : :
690 SerProSerThrSerGlyMetSerThrLeuThrSerGluProSerProSe 706
576 ACACACATGAGTATTTACGTTTCGGTCAGCGGCTCAATACCAACAGCGCG 527
||| : : : : : ||||| : : : : :
706 rSerThrGlnSerSerGlyAlaGlnSerThrLeuThrThrProSerProA 723
526 GTTGAAGTCCTTGGCGGCTTCTTTCATGATGACCGTAGGTCGGCAGCC 477
: : : : : ||||| : : : : : : : : : : : : : : :
723 snProSerGlnSerThrSerSerLeuGluSerSerThrSerGlyAlaThr 739
476 ACGGATGCTGCTCATCGATTCGAGAGATGCGGAACGGCTCGGCATC 427
: : : : : ||| : : : : : : : : : : : : : : :
740 ThrSerSerGlySerAlaGlyThrThrMetThrSerProSerGlnSerSe 756
426 TACGCGAGGATTTTCTGAACGAGCGGCTGCGAAGCGCGATCCATAAGC 377
:
756 rSerValGlySerSerGlnGlySerThrSerProAlaAla..... 769
376 CTGATTGAATCAGGTTCGCGCGCACTTTTCGCTGCTCAATTTTCCACG 327
:
770SerThr 771
326 GCTTCAGTACGTAGCGTTTCGAACGATTCGTTGCTTCCCTTCAACGCG 277
: : : : : ||||| : : : : : : : : : : : : : : :
772 ThrSerGlyGluMetThrSerGlnGlySerThrGlnThrProGlySerSe 788
276 AATCAGCATGACTGAAGTACGCGCTTTTCGCGACGGTGAATAGCGGGA 227
:
788 rValSerThrSer.....AlaAlaI 795
226 TTTTTCCTCAAGCC.....GCGCAGTA 204
||||| : : : : : ||||| : : : : : : : : : : : : : : :
795 leLeuThrSerThrGlnGlnSerValSerThrAsnSerProGlySerThr 811
203 AATPACTACGCGGATCTTTTGTCTTCAAAACAGCATTGCGCTTTT 154
||| ||||| : : : : : : : : : : : : : : : :
812 ValThrArgProSerThrValSerGlySerThrSerSerGlySerThrVa 828
153 GACGCTTACCTTCCTGATTTTCATCAGAGGCGCGCATCCGACATATT 104
:
828 lThrValGlySerThrGluAlaSerThrSerGlySerValAlaSerS 845
103 CTTTCGCAAGCAACGACTTCGTAATGCGCGGCGCTCATATAATGACT 54
||||| : : : : : ||||| : : : : : : : : : : : : : : :
845 erSerProAlaProSerThrSerGlnAsnProAsnProSer.....Thr 859
53 TGCTCCGGT 45
|||||
860 SerSerGly 862

seq_name: sp_bacteria:Q9RFF2

seq_documentation_block:
ID Q9RFF2 PRELIMINARY; PRT; 569 AA.
AC Q9RFF2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE UREC.
OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1063;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2.4.1;
RA Choudhary M., Kaplan S.;
RT "DNA sequence analysis of the photosynthesis region of Rhodobacter
RT sphaeroides 2.4.1."; 0:0-2000).
RL Nucleic Acids Res. 0:0-2000).


```

327 uAspPro.....ArgGlyC 332
613 GCCAATATCGAAACACATGAATTTGGCGCGCCGATCTCGCGGCTTGAG 662
332 ysglyLeuArgArg.Glu.....ProHisProAlaArgLysTh 344
663 TGGCAGCAGACATTCATTTCATCGAGCAGTCGCGCGCAATAAACCCTGT 712
344 rLeuAlaAlaGluAspLeuHisAspMetGlyAla.....P 357
713 GGACATCAATATCAGACACGTGATGCTATCGGACGCTTTGTTCTGTAACA 762
357 heSerValIleSerSerAspSerGlnAlaMetGlyArgVal..... 370
763 GCGCGTCTGAATACCGAGCGGTGTCCTTGGCGCGCTGCAAGTCAA 812
371 GlyGluValIleThrArgThrTrpGlnThrAlaHisLysMetLysValG 387
813 CAACCGCGCCTCTTGGGTACCGCTTTGGGTGCGAAGGTGCTCAACTTA 862
387 nArgGlyArgLeuAlaGluGluThr...GlyAlaAsnAspAsnGlnArgV 403
863 CCGCGCGCGAATGTTGTCACCGGACAAACCGCGTATTCGGTTCGGTA 912
403 ai.....ArgArgTyrlleAlaLysTyThr 411
913 TTGAACGCTGCGATTGCAAGCGCGCATGATTATTGCGACGCTACCA 962
412 IleAsnProAlaIleAlaHisGlyLeuSerArgHisIleGlySer..... 426
963 CAATCAGATTTCGTTATCGAAGAGCGCGCAGCAAGAGCTGTTCGGCT 1012
427 .....ValGluGluGlyLysArgAlaAspLeuValLeuT 438
1013 GGTTGCGCGCGAGCGG 1029
438 rp.....GlnPro 440

seq_name: sp_invertebrate:Q9U0V2
seq_documentation_block:
ID Q9U0V2 PRELIMINARY; PRT: 2207 AA.
AC Q9U0V2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE POSSIBLE MUS308 HOMOLOG (FRAGMENT).
OS Leishmania major.
OC Eukaryota; Eudlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RC STRAIN=FRIEDLIN;
RA Masuy D., Purnelle B., Goffeau A., Ivens A.C., Lawson D., Quail M.,
RA Rajandream M.A., Barrell B.G. to the EMBL/GenBank/DBJ databases.
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=FRIEDLIN;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
DR EMBL; AL122012; CAB58415.1;
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase.C.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; Helicase.C; 1.
DR SMART; SM00490; HELICC; 1.
KW ATP-binding; Helicase.
```

```

FT NON_TER 2207 2207
SQ SEQUENCE 2207 AA; 229262 MW; E461565DA0017B0C CRC64;

alignment_scores:
  Quality: 108.00      Length: 452
  Ratio: 0.565        Gaps: 23
  Percent Similarity: 42.257      Percent Identity: 22.788

alignment_block:
US-09-303-518D-131/rev x Q9U0V2 ..
Align seg 1/1 to: Q9U0V2 from: 1 to: 2207

1232 CAACCCAAAGCCGTCGCGCTGCTGCGTATCGCGACGATTAAATCGCGCAA 1183
813 GlnProArgLeuCysGlnLeuSerValAlaThrThr..... 824
1182 AAGCAAGGTAGGAGGATGTCACACGGCATTACGGCC.....T 1145
825 .....ThrGlyGlnAlaSerAlaGlySerAlaArgAsnValAlaAlaS 839
1144 CATAAGTCCGATCGGTACCATCGCGCGGTGCGCGCTGACGGCTGTC 1095
839 erValAlaGluProAlaThrProValSerSerProSerGluAla... 854
1094 GTCAACTTCAAGAGTTGTTTGTAGGAAATGGCCGAGAGTGGTCGCGT 1045
854 ..... 854
1044 GATGGAGTATTTGTCGCGCTGCGCGCACCCAGCCGACACGCTCTTTCG 995
855 .....CysAlaThrThrHisPro...ProAspLeuT 864
994 TCGCGCCTTCTTCGATAACGGAATCTGATTGTGTAGCTAGCTCCCAATAA 945
864 rpLeuProSerLeuProThrGluHis..... 872
944 TCATGCGCGCCTTGTGCAATCGCACCGCTTCAATACCGAACCGGAATCAC 895
873 ValSerProProcys.....AspProLeuLeuMe 882
894 GCGGTGTCCGCGCTCAACCAATTCGCGCGGTAAGTTGAGACACCTTCG 845
882 targValSerSerGlnGlyGly..... 889
844 CACCCAAACGGTACGCAAGAGCGCGGTGTTGACTTGCAGGCGCGCC 795
890 .....LysAlaGluGly..... 893
794 AAGGCAACACACGCTCGCTGCTATTCAGACGCGCTGTTCAGAACACGTC 745
894 .....CysAlaAlaGlnSerValPheargMetGlyValValAlaProThrPr 909
744 GATACATACGCTCTTCGATAATTGATGGTCCACACGCGTTTATTTCGCG 695
909 othrSerLeuGlySer.....AspValLeuSerSera 920
694 CGACTGCTCGATGAAATGAATGTGCGTGCCTCACCTCAAGCCGCGAGGATGC 645
920 laThrGly.....ValSerAlaLeuSerAlaAlaAsnAla 931
644 GGGCGCCAAATTCATGTGTTTCGATATTGCA..... 612
932 AlaProProLeuSerAspLeuHisValThrAlaLeuProTyrLeuAlaTh 948
611 .GCATTTTCAGCGCAGCTCTGCGCTGCT..... 582
948 rAlaAlaAlaGlyGlySerGlyAlaProAlaValProAlaArgProGlyA 965
581 .....GCTTTACACACATGGATTTCAGTTTCGTCAGCGCGCTC 543
965 rgThrCysPheThrLeuHisSer.....AlaAlaArgThrThr 977
```

```
542  ATACCAACAGCCGCGTTGAAGTCTTCGGGGTCTTTG..... 501
978  GlyThr.....LeuSerSerAlaAlaAlaThrSerThrG1 990
500  ....ATGATGACCGTAGGTCGGCAGCGCGATTGGTGTCCATCG 458
990  uGluProThrThrAlaAlaAlaSerAlaSerThrGlyAlaValProProA 1007
457  CATTGACGAAGATGCGAAGCGCTCGGCATCTACGGCAGGATTTGCTG 408
1007  laLeuThrGluAlaSerfyrLeuSerAlaGluLysCys.....Phephe 1021
407  ACGGACGGTCCGAAGCGCAGTGCATAGCTGATTGAATCAGATTGCG 358
1022  ArgGlyGlyThrAlaGluLeuGlnGlnLeuProGly..... 1033
357  GCGCACATTTTTCGCTGCTCAATTTTCCAGCGCTTCAGGTAGCTAGCGTT 308
1034  .....HisGlyAlaAlaSerProGlnGlyThr.....G 1043
307  CGAAGCTGATTCGCTGCTGCTTCAACGGCAATCAGCACTGCTGAAGT 258
1043  luGlyLeuMetValSerLeuAspGlyValAlaSerAlaThrProSerAsp 1059
257  ACGCGCTTTTCGCGCGGTGAATA..... 234
1060  ThrArgLysMetProSerSerMetGluSerArgHisThrThrGlyCysAl 1076
233  .....GCGGCGATTTG..... 222
1076  aGlnProValIleAlaAsnAlaGlnProAlaLeuLeuArgProValAsps 1093
221  .....CCTGAAGCGCGCAGTAATACTACCGCGGATCTTTTGTCT 177
1093  erThrProSerThrValValMetHisArgThrProSerThrLeuThrSer 1109
176  TCAACAGCAGCTGG.....CCTTTTTCGCGGCTTCACC 142
1110  GluGluProAlaTrpAsnArgSerAlaProPheProGlnProTrpPr 1126
141  TTCCTTTGATTTTCATCGAGGGCGCGATCGACATATTTCTCGCAAGCA 92
1126  oSerLeu.....SerAspGlyGlnProProAlaLeuLeuThrValAlaA 1141
91  ACGCGACTTCGGTAATGCGCGGCGCTCATATGACTTCTCCGCTCG 42
1141  laValThrProLeuValGly.....CysMetGlyAla 1151
41  CCGCGG 36
1152  GluAla 1153
seq_name: sp_invertebrate:Q9VXM5
seq_documentation_block:
ID Q9VXM5 PRELIMINARY; PRT; 2406 AA.
AC Q9VXM5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE CG9056 PROTEIN.
GN CG9056.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Wortman J.R., Yeaman M.D., Zhang Q., Chen L.X.,
Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
Ballew K.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Durlin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
Foster K.C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimball M.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Spieler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J., Yao Q.A.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT The genome sequence of Drosophila melanogaster.;
RL Science 287:2185-2195(2000).
DR EMBL; AF003500; AAF48533.1; -.
DR FlyBase; FBgn0030702; CG9056.
SQ SEQUENCE 2406 AA; 248803 MW; 740EAD78D684F76C CRC64;
```

```
1341 erGlyAlaSerGlnProValValSerValProSerAlaAsnSerThr 1357
::: ||| |||:::||||| |||
1082 AGTTTGTGTTTGTAGAAATGCGGAGACTGGTG.....CGCGTCAT 1042
:::||||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
1358 AlaLeuSerMetSerAsn....ProHisIleSerHisSerHisValPr 1373
:::||||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
1041 GGAGTATTGTCGGC.....TGCGGCGCAACCC 1013
||| ||||| |||::: |||::: |||::: |||::: |||::: |||::: |||
1373 oAlaTyAlaSerGlyAlaPheSerSerSerAlaAlaAlaGlyThrSerT 1390
1012 AGCCGAACAGCTCTTG.....CTGCGGCTTCTTCGATTAACGGAATC 969
|||||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
1390 hrProAsnSerGlyLeuSerThrLeuAlaValThrSerLeuSerThr... 1405
968 TCATTGTGCTAGCTCCCAATAATCATGCGCGCTGTGTGCAATCGCAC 919
1406 .....SerAlaAlaPro.....GlnPr 1411
918 GTTCAATACCAACGGAATACAGCGG...TTGTCCGCGTCAACCAATT 872
|::: |||::: |||::: |||::: |||::: |||::: |||::: |||
1411 ohSerHisPheProGlnSerThrGlnMetLeuProGlnSerGlyAsnPr 1428
871 CGCGCGGTAAGTTGACACACCTTCGCACCAACACGGTACGCAAGG 822
1428 heSerSerValSer..... 1432
821 CGCGGTTGTTGACTTGCAGCGCGCCCAAGCAACACGCGCTCGGTATT 772
|||||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
1433 ...HisLeuLeuThrHisProMetSerSerGlnAsnGlnPrometVa 1448
771 CAGACGGCTTTACGACAAACCTCGTAGCAATCACGCTCTTGATAAT 722
1448 larg..... 1449
721 TGATGTCCACACAGCTTTTATTCGCGCGACTGCTCGATGAATGATG 672
1450 .....CysGlySerThrLeuSerGlnSerSerAla..... 1460
671 TGCCTGCCACTCAAGCGCGGAGGATGGCGCGCCAAATTCATGTGT.. 624
|||||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
1461 .....AlaAlaThrAlaProProSerAlaAlaAla 1471
623 .....TCGATATTGCCAG 611
1471 aValSerAsnPheThrProSerValLeuAlaValGlnSerLeuThrA 1488
610 CATTTTCAGACGCGACTCTGCGGCTGCTGCTTACACACATGGATTTA 561
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
1488 laValThrSerSerSerSerProSerThrLeuSerSerSerValle 1504
560 CGTTCGGTC.....AGCGGCTCAATACCAACAGGCC 529
:::||||| |||::: |||::: |||::: |||::: |||::: |||::: |||
1505 GlnLysValIleSerProLysGlyGluSerProCysAsnLysAspArgAs 1521
528 GCGTTTGAAGCTTTCGCGGCTTCTTTGATGATGACCGTAGGTCGGCAG 479
|||||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
1521 pSerSerTySerSerProAlaAsnAlaValValThrThrCysAlaPro 1538
478 CCAGCGGATTTGTTCCATCGCATGACGAAGATGCGCAACGCGTCGGCA 429
:::||||| |||::: |||::: |||::: |||::: |||::: |||::: |||
1538 hrThrProIleValSer.....SerGlySerAla 1547
428 TCACGCGAGGATTTGTGACGAGCGGGTTCGAAGCGAGTCATCAATAA 379
::: ||| |||
1548 ArgProThrProProLeuSerAsn..... 1555
378 GCCTGATTGAATCAGGTTGGCGGCACTTTTCGCTGCTCAATTTGCCA 329
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
1556 .....CysThrSerMetGlyIleGlyMetValA 1565
328 CGCGTTCAGTACGTACGCTCG.....AATCGATTTCG.....TCG 291
:::||||| |||::: |||::: |||::: |||::: |||::: |||::: |||
```

```
1565 snAlaAlaSerThrAlaArgSerSerCysAsnAlaIleSerProLeuSer 1581
290 TTGCCTTCAACGGCAATACGACTGACTGAAGTACGCGCTTTTCGCCACG 241
:::||||::: |||
1582 IleProAlaThr..... 1585
240 GTGAATAGCGCGATTTTGCTGAAGCCGCGCAGTAATAATACACGCCG 191
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
1586 .....AlaGlyIleHisValSerAla.....ThrAsnProS 1596
190 GATTCTTTTGTCTCAACAGCAGCTTGGCTTTTGTGACGGCTTCACCT 141
:::||||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
1596 erPhe.....GlnSerSerSerTyPheProThrProLeuAlaProPro 1610
140 TCCTTGATTTCATCGAGGGCGCATGCCGACATATTTCTTCGCCAAGCAA 91
|||||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
1611 Pro.....SerSerProSerPr 1616
90 CGCGACTTCGGTAATGGCC 72
1616 oAlaThrSerSerAlaAla 1622
seq_name: sp_human:Q9H195
seq_documentation_block:
ID Q9H195 PRELIMINARY; PRT; 901 AA.
AC Q9H195;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MUC3B MUCIN (FRAGMENT).
GN MUC3B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Pratt W.S., Crawley S., Hicks J., Ho J., Nash M., Kim Y.S., Gum J.R.,
RT Swallow D.M.
RT "Multiple transcripts of MUC3: Evidence for two genes, MUC3A and
RT MUC3B."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ291390; CAC19572.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000082; SEA.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS00024; SEA; 1.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 901 AA; 96192 MW; F98FC817494ECD99 CRC64;
```

alignment_scores:

Quality: 107.00 Length: 440
Ratio: 0.532 Gaps: 16
Percent Similarity: 45.682 Percent Identity: 20.909

alignment_block:

US-09-303-518d-131/rev x Q9H195 ..

Align seg 1/1 to: Q9H195 from: 1 to: 901

```
1292 TCGTATTTTCGGCGGCGAGCAAGCTGCACAAAGCGAGGTCTTCTTCGTC 1243
|||||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
6 SerIleIleThrThrGluThrThrSerHisSerThrProSerPheSerSe 22
1242 CAATTCCCAAGCAACCAAGCGCTGCGCTGCTCGGTATCGGCGACGATTA 1193
|||||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
22 rSerThrIleHisSer.....ThrValSerSerSerThrT 34
1192 AATCCGCGCAAAAGCAAGGTAGGAGATGTCCAAACGGCATTTACGCGTCA 1143
```

```

34 hrAlaIleThrSerProPheThrAlaGluThrGlyValThrSerThr 50
1142 TAAGTCCGATCGTACCATCGCGGCGCGCGGTTGACGGCTGCTG 1093
51 ProSerSerProSerSerLeuSerThrAspIleProThrThrSer.... 65
1092 GAACCTGAAGAGTTGTTTTAGGAATAGCCGAGAGTGGTCCGCTGA 1043
66 .....LeuArgThrL 69
1042 TGGAGTATTGTCGGCTCGCGCAACCCAGCGCAACCGGAAATCAGCG 893
69 euThrProLeuSerLeuSerThrSerThrSerLeuThrThrThrThr 85
992 CGGCTCTTCGATACGGAATCTGATGTGTAGCTGCCAATAATC 943
86 LeuProSerIleProThrAspIle.....Se 94
942 ATGCGCGCTTGTGCAATCGCACCTTCAATACGCAACCGGAAATCAGCG 893
94 rSerLeuProThrProIleHisSerSerProSerProSerIleGlns 111
892 GGTGTCCGCTCA.....ACCAATTCGCGCGGTAAGTTGA 855
111 erThrGluThrSerSerLeuValGlyThrThrSerProThrMetSer... 126
854 GACACCTTCGACCCAAAACGGTACCAAGAGCGCGGTTTGTGACTG 805
127 .....ThrValArg..... 129
804 CAGCGCGCCCAAGGCAACCGCGCTCGGTATTCAGACGCGCTGTACGA 755
130 .....AlaThrLeuArgSerThrGluAsnThrProIleSerS 142
754 AC...AAACGTCGATAGCAATCATCTTGTATATGATGTCACACG 708
142 erPheSerThrSerIleValValThrPro..... 151
707 GTTTATTTCGCGCGCTGCTCGATGAATGATGTCGTCACCTCAA 658
152 .....GluThrProThr.....GlnAlaProValLe 162
657 GCGCGGAGGATCGGCGCGCAATTCATGTTTCGATATTTGGCAGCAT 608
162 uMetSerAlaThrGlyThrGlnThrSerProValProThrThrValThrP 179
607 TTCACAGCGCAGCTGTCGCGCTGCTGCTTACACACATGGATTACGT 558
179 heGlySerMetAspSerSerThrSerThrLeuHisThrLeuThrProSer 195
557 TCGGTC.....AGGCGGCTCAATACCAACAGCGCGGTTTGAAGCTTC 514
196 ThrAlaLeuSerIleMetSerThrSerGlnPheProIleProSerTh 212
513 GCGCGCTCTTGTATGATGACCGTAGGTCGCGCAGCGGATGGTGT 464
212 rHisSerSerThrLeuGlnThrThrProSerIleProSerLeuGlnThr 229
463 CCATCGATTTCAG.....AAGATGGCGAAGCGGCTCG 432
229 erLeuThrSerThrSerGluPheThrThrGluSerPheThrArgGlySer 245
431 GCATCTACGCGAGGATTTGCTGAACGAGCGGTCGCGAAGCGCAGTCCA 382
246 ThrSerThrAsnAlaIleLeu..... 252
381 TAAGCCTGATTGAATCAGGTTGGCGGCACCTTTTCGCTCTCAATTTG 332
253 .....ThrSerPheSerThrIleIleTrps 261
331 CCAGCGCTTACGTTAGCTAGGTTTCGAACCTGATTTCGTCGTCGCTTCA 282

```

```

261 erSerThrPro.....ThrIleIleMetSerSerSerProSer 273
281 ACGCATACGACTGACTGACTGACGCGCTTTTCGCCACGGTGAATAGC 232
274 SerAlaSerIleThr.....Pr 279
231 GGGGATTTTGCCTCAAGCCGCGCAGTAAATACAGCCGCGGATTTT 182
279 oValPheAlaThrThrIleHisSerValProSerSerProTyIlePheS 296
181 TGTCTTCAACAGACTTGGCCCTTTTGTGACGCTTCACCTTCTTGTATT 132
296 erThrGluAsnValGlySerAlaSerIleThrAlaPheProSerLeu... 311
131 TTCATCAGGGGCGCATGCGCATATTTTCGCAAGCAACGCGACT.. 84
312 .....SerSerSerSerThrThrSerThrSerProThrSerSerLe 326
83 .....TCGGTAATGCGCGCGGCT 65
326 uThrThrAlaLeuThrGluIleThrProPheSerThrIleSerLeuProS 343
64 CATAAATGACTTGCCTCGGT 45
343 erThrThrProCysProGly 349
seq_name: sp_rodent:Q9QY35
seq_documentation_block:
ID Q9QY35 PRELIMINARY; PRT; 1979 AA.
AC Q9QY35;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HOST CELL FACTOR C1 (FRAGMENT).
GN ABCD1 OR HCFC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Platzer M., Brenner V., Reichwald K., Wiehe T., Oksche A.,
RA Rosenthal A.;
RT "Comparative sequence analysis of the mouse L1cam locus and the
RL corresponding region of human Xq28.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AFI33093; AAF22156.1; -.
DR MGD; MGI:1349215; Abcd1.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR001798; Kelch.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF01344; Kelch; 3.
DR SMART; SM00060; FN3; 1.
FT NON_TER 1
FT SEQUENCE 1979 AA; 203436 MW; ABE998C5DDE688A9 CRC64;
SQ
alignment_scores:
Quality: 107.00 Length: 464
Ratio: 0.510 Gaps: 24
Percent Similarity: 45.259 Percent Identity: 22.198
alignment_block:
US-09-303-518d-131/rev x Q9QY35 ..
Align seg 1/1 to: Q9QY35 from: 1 to: 1979
1118 CGGTGCGCGCGCTTACGCGCTGC.....GTGAACCTGAAGAG 1081
297 LysProProProAlaArgValGlnLeuValArgAlaAsnThrAsnSe 313
1080 TTTGTTTTTAGGAAATGCGCGAGAGTGGTGGCGGTGATGAGTATTG. 1032

```

```
|||||
313 rLeuGluVal...SerTrpGlyAlaValAlaThrAlaAspSerTyrLeuL 329
1031 .....TCCGGCTGGGGCGCAACCCAG 1011
329 euGlnLeuGlnLysTyrAspIleProAlaThrAlaAlaThrAlaThrSer 345
1010 CGAACACGCTCTTCTCGCGCTCTTCGATAACGGAAATCTGATTGTG 961
346 ProThr.....ProAsnProValProSerValProAla... 356
960 GTAGCGTCCCAATAATCATCGCGCTTGTGCAATCCGACCG..... 918
357 .....AsnProProLysSerProAlaProAlaAlaAlaProAlaValG 372
917 .....TTCATACCGAACCGGAATCACCGGTTG 888
372 lnProLeuThrGlnValGlyIleThrLeuValProGlnAlaAlaThrAla 388
887 TCCGGGTCAACAATTCG.....CCGGCGGTAACTTGAGACAC 850
389 ProProSerThrThrIleGlnValLeuProThrValProGlySerSe 405
849 CTTCGACCCCAAAAGGTACCAAGAGCGCGGT.....TTGTGA 809
405 rIleSerValProThrAlaAlaArgThrGlnGlyValProAlaValLeuL 422
808 CTTCAGCGCCCAAGCAACCAACCGCGCTG.....GTATTCAGACGG 765
422 ysValThrGlyProGlnAlaThrThrGlyThrProLeuValThrMetArg 438
764 CCTGTTAG.....ACAAAGTCCGATAGCAATCATCGTCTTGATAATT 721
439 ProAlaSerGlnAlaGlyLysAlaProValThrValThrSer.....Le 453
720 GATGTCACACGGTTTATTCGCGCGAGGATGCGGG..... 642
453 uProAlaSerValArgMetValValProThrGlnSerAlaGlnGlyThrV 470
670 CGGTGCCACTCAAGCGCGAGGATGCGGG..... 642
470 alIleGlySerAsnProGlnMetSerGlyMetAlaAlaLeuAlaAlaAla 486
641 .....CGCCAAATCATGTGTTTCGATATGCG 613
487 AlaAlaAlaThrGlnLysIleProProSerSerAlaProThrValLeuSe 503
612 A.....GCATTTTCAGACGGCA 596
503 rValProAlaGlyThrThrIleValLysThrValAlaValThrProGlyT 520
595 CGTCTGCGCTGCTGCTTTACACACATGGATTTTACGTTCCGTCAGCGCG 546
520 hrThrThrLeuProAlaThrValLysValAlaSerSerProValMetVal 536
545 CTCATACCAACAGCGCGCTTGAAGTCTCGCGCGCTTCTTTGATGAT 496
537 SerAsnProAlaThrArgMetLeuLysThrAlaAlaGlnValGlyTh 553
495 GACCGTAGGTGCGCGAGCC..... 477
553 rSerValSerSerAlaAlaAsnThrSerThrArgProIleIleThrValH 570
476 .....ACCGGATTTGTTCCATCGCA.....TTGACGAAG 447
570 isLysSerGlyThrValThrValAlaGlnGlnAlaGlnValValThrThr 586
446 ATGGCGAAGCGCTCGCATCTAG..... 423
587 valValGlyGlyValThrLysThrIleThrLeuValLysSerProIleSe 603
422 .....GCAGGATTTTGTGAAC...GGACGGGTGCGAAGCG 389
..... ||| ||| ||| ||| |||
```

```
603 rValProGlyGlySerAlaLeuIleSerAsnLeuGlyLysValMetSerV 620
388 CAGTCCAT...AAGCCTGATTGAATCAGGTTGCGGCGCACTTTTTCGCTG 342
620 alValGlnThrLysPro.....ValGlnThrSerAlaVal 631
341 CTCATTTTTCGACGCGT..... 324
632 ThrGlnGlnAlaSerThrGlyProValThrGlnIleIleGlnThrLysG 648
323 .....TCAGGTACGTAGCGTTCG.....AACTCGATTTCGTCTG 290
648 yProLeuProAlaGlyThrIleLeuLysLeuValThrSerAlaAspGlyL 665
289 TGCTTCACGGCAATCAGCACTGACTGAAGTACGCGCTTTTGGCCACGG 240
665 ysProThrThrIleIleThr..... 672
239 TGAATAGCGCGATTTCGCTGAAGCGCGGCAGTAATACTACGCGCGG 190
673 .....ThrGlnAlaSerGlyAlaGlyThrLysPro... 682
189 ATTCCTTTTCTCTCAACAGCACTTGGCTTTTTCAGCGCTTCACCTT 140
683 .ThrIleLeuGlyLeSerSerValSerProSerThrThrLysProGlyT 599
139 CCTTGATTTTCAGAGGGCGCATCGGACATATTTCTTCGCCAAGCAAC 90
699 hrThrThrIleIleLys.....ThrIleProMetSer 709
89 GCGACTTCGTAATGCGCGCGCTCATATAATCAGTTGCTCC 48
710 AlalleIleThrGlnAlaGlyAlaThrGlyValThrSerSer 723
seq_name: sp_human:Q14395
seq_documentation_block:
ID Q14395 PRELIMINARY; PRT; 505 AA.
AC Q14395;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MUCIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=STOMACH;
RX MEDLINE=93194895; PubMed=7680650;
RA Toribara N.W., Robertson A.M., Ho S.B., Kuo W.L., Gum E., Hicks J.W.,
RA Gum J.R. Jr., Byrd J.C., Siddiki B., Kim Y.S.;
RT "Human gastric mucin. Identification of a unique species by expression
RT cloning.";
RL J. Biol. Chem. 268:5879-5885(1993).
DR EXBL: L07518; AAB61945.1; -.
FT NON_TER 505
FT NON_TER 505
SQ SEQUENCE 505 AA; 51967 MW; 45D7CB1E6050E07D CRC64;
```

```
alignment_scores:
  Quality: 106.50      Length: 492
  Ratio: 0.475        Gaps: 21
  Percent Similarity: 45.528  Percent Identity: 20.325
```

```
alignment_block:
US-09-303-518d-131/rev x Q14395 ..
```

```
Align seg 1/1 to: Q14395 from: 1 to: 505
```

```
1340 CCTTCCTCTCAATGGTTCAGCACTTTCCAGCACTTTGCGCAACAGCGCGGTATTC 1291
```

```

339 SerThrSerSerValThrProSerThr...HisThrValIleThrProth 354
555 GGTCAGCGGCTCAATACCAACAGCGCGCTTTGAAGTCTTCGGCGCTT 506
354 rHisAlaGlnMetSerThrSer..... 361
505 CTUTGATGATGACGCTAGGCTCGCGACGCCAGGATTTGGTGTCCATCGCA 456
362 .....AlaSerIleHisSerThrProThrGlyThrValProLeu 375
455 TTGACGAGAGTGGCGAACGCTCGGCATCTACGCCAGGAGTTTGTCTGAA 406
376 ThrThrArgMetProThrGlySerThrArgThr...GlyProMetTh 391
405 CGGACGGGTGCGAAGCGCAGTCCATAAGCCTGATTGAATCAGGTTCGGC 356
391 rGlyThrIle.....IleGlnThrSerL 399
355 GCACCTTTTTCGCTCTCAATTTTCCAGCGCTTCAGGTACGTAGCGTTTCG 306
399 ysAlaHisAsnSerPheSerThrAlalysThrSerThrSerLeuHisSer 415
305 AACTCGATTTCGTCG...TTGCCTTCAACGCAATCACGACTGACTGAAG 259
416 HisAlaSerSerThrHisHisProGluThrThrProThrSer.....Th 430
258 TACGGCTTTTCGCGACGGTGAATAGCGGGGATTTTCCTTCAAGCGCGC 209
430 rThrAsnIleThrProLys.....SerThrSerAlaGlyT 442
208 CAGTAATACTACGCGCGGATCTTTTGTCTTCAACAGCAGCTTGGCCT 159
442 hrSerThrProValAlaHisThrThrLeuAlaThrSerSerArgLeuPro 458
158 .....TTTTGACGGCTTCACCTTCCTTGATTTTCATCGAGGGCGCAT 115
459 ThrThrPheThrThrHisPhePro..... 466
114 GCGGACATATTTTGGCAACAGCGGACTTCGGTAAATGCGCGGCGCT 65
467 .ProThrGlySerSerHisValSerSerThrGlyProMetThrAlaThrS 483
64 CATAAATGACTTCGCTCGCTGCCC 39
483 erSerGlnThrThrThrHisPro 491

seq_name: sp_bacteriap:Q98K15

seq_documentation_block:
ID Q98K15 PRELIMINARY; PRT; 515 AA.
AC Q98K15;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DE ATP-BINDING PROTEIN OF RIBOSE ABC TRANSPORTER.
GN MLL1680.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Igesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto M.,
RA Mochizuki Y., Yamada M., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
DR EMBL; AP002998; BAB8999.1;

```

DR InterPro: IPR003593; AAA.
 DR InterPro: IPR003439; ABC_transport.
 DR InterPro: IPR001687; ATP_GTP_A.
 DR Pfam: PF00005; ABC_tran; 2.
 DR SMART: SM00382; AAA; 1.
 KW ATP-binding; Complete proteome.
 SQ SEQUENCE 515 AA; 55185 MW; 882AC3E521F5A1B6 CRC64;

alignment_scores:
 Quality: 106.50 Length: 483
 Ratio: 0.498 Gaps: 17
 Percent Similarity: 44.306 Percent Identity: 18.841

alignment_block:
 US-09-303-518D-131 x Q98K15 ..
 Align seg 1/1 to: Q98K15 from: 1 to: 515

```

88 GCGTTGCTGGGGAAGAATATGTCGGCATGCGCCCTCGATGAATCAA 137
|||||
37 AlarileuGlyGluAsnGlyAlaGlyLysSerThrLeuIleLysValMe 53
|||||
138 GGAAGGT...GAAGCCGTCAAAAAGCCAAAGTCTGTTTGAAGACAAAA 184
|||||
53 talaglyValValAlaProAspGluGlyArgMetThrLeuAspGlyArgG 70
|||||
185 AGAATCGGCGGTAGTATTTACTGCGCGCGCTTCAGGCAAAATCGCGCT 234
|||||
70 lu.....ValThrPheLeuSerProAlaAlaAlaAsnGlnAlaGly 83
|||||
235 ATT..... 237
|||||
84 lleValcysIlePheGlnGluLeuSerLeuValProAspLeuSerValAl 100
|||||
238 .....C 238
|||||
100 aaspAsnIleValIleSerAspProLysArgPheGlyMetIleAspA 117
|||||
239 ACCGTGGCGAAAGCGGTACTTACGTAGTCAGTGTGCGGTGAAGGC 288
|||||
117 rglYsAlaGlnArgValAlaGluAlaGluAlaAlaAlaAlaGlyAla 133
|||||
289 AAGCAC.....GAATCGAGTTCGAACGCTA 314
|||||
134 AlaAspIleHisProLeuAlaLeuValLysAspLeuProLeuSerArg 150
|||||
315 CGTACTGAAGCGCTGGCAAAATTTAGCAGCAAAAAAGTCCGCGCAACC 364
|||||
150 gGlnMetValGluIleAlaLysAlaLeuAlaArgLysProArgIleLeu 167
|||||
365 TGATTCAA.....TCAGCTTATGACACTGCGCTTCGCACCGCTCGG 405
|||||
167 leLeuAspGluAlaThrSerAlaLeuThrAlaAlaAspValSerLysIle 183
|||||
406 TTCAGCAAAATCCCTGCCGTAGATGCGAGCGCTTCGCCATCTTCGTC 455
|||||
184 PheGlyValLeuLysArgLeuArgSerGluGlyLeuAlaLeuLeuTyr 200
|||||
456 TCGCATGGACACCAATCCGCTGGCT.....GCCAGCCCTACGGTCATCA 499
|||||
200 eSerHisArgMetAsnGluIleAlaGluLeuAlaAspGlnCysThrValP 217
|||||
500 TCAAGAAGACC.....GCCAGACTTCAAAACGGCGCTGTGTGATTG 543
|||||
217 heArgAsnGlyArgAsnValAlaLysTyrLysAlaGlySerLysSerAsp 233
|||||
544 AGCCGCTGACCAACGTAATAATC.....CATGTGTGTAA 578
|||||
234 AsnGluValValGluLeuMetIleGlyArgGluTyrSerHisIlePheP 250
|||||
579 ACAGCAGCGCGAGCGTGGCGCTCTGAAAATGCTGCCAATATCGAAAC 628
|||||

```

```

250 oAlaLysProAlaAlaValProAlaThrAlaAlaProArgLeuGlu.... 265
629 ATGAATTGGCGCGCGCATCTGCGGGTGTAGTGGCAGCACCATTCAT 678
265 ..... 265
679 TTCATCGAGCCAGTCGGCGCGAATAAACCGTGTGGACCAATATATCA 728
266 .....AlaArgLysLeuSerTrpThrAspArgLeuHi 276
729 AGACGTGATTGTCGGAGCTTTCGTCTAACAGCCGCTCTGAATACGG 778
276 sasPileSerLeuThrValArg.....AlaG 285
779 AGCGCGTGTGCTTGGCGCGCTGCAAGTCAACAACCGCGCCTCTTG 828
285 lyGluValValGlyLeuGlyGlyLeuAspGlyGlnGlnArgGluLeu 301
829 CGTACCGTGTGGTGCAGAGGTGTCTCAACTTACCGCGCGGGAATGTGT 878
302 LeuLeuAlaPhePheGlyValLeuArgGlyLeuSerGlyGluValLeu 318
879 TGACCGCGCAACCGCGTTCGGTTCGGTTCGATTGAACGGTGCATGG 928
318 easpGlyLys.....ProValAlaIleGlySerProA 329
929 CACAAGCGCGCATGATTATTGGGACGCTACCAACATCAGATTTCCGTT 978
329 laLysAlaArgLysGlnaspGlyIleGly.....MetAlaLeu 340
979 ATCGAAGAGCGCGCAGCAAGAGCTGTGCGCTGGTTCGCGCGCAGCC 1028
341 lleProGluAspArgLysThrGlu.....GlyLeuMetLeuProMetH 355
1029 GCACAAATCTCCATCAGCGCGCACCCTCTCGGCCATTTCTCTAAACA 1078
355 rValArgGluAsnLeuSerPheAlaAlaLeuAspArgLeuSerLysGly 372
1079 AACTCTTCAAGTTCACGACA..... 1098
372 lyIleAspArgAlaAlaGluGlnArgLeuIleAspMetValGly 388
1099 .....GCCGTCAA 1106
389 LeuLeuAlaIleLysThrAlaGlyLeuAspIleProValGlyAlaLeu 405
1107 CGCGCGCGCGCATGTTACCGTACCGTACCGCATTTATGAGCGCGTAATGC 1156
405 rGlyGlyAsnGlnGlnLysValValIleAlaLysTrpLeuMetArgGlnP 422
1157 CGTTGGACATCTG.....CCTACC..... 1176
422 roArgIleLeuLeuAsnAspProThrArgGlyIleAspValGlyThr 438
1177 .....TTGCTTTTCGCGATTTA..... 1194
439 LysGlnGluLeuTyrGlnLeuMetArgLysLeuAlaAspAlaGlyAla 455
1195 .....ATCGTCGGCGATACCGACGCGCGCGCTTTCGGTGTGTGAA 1239
455 aIleLeuPheTyrSerThrAspTyrAspGluLeuIleGlyCysCysAsp 471

```

seq_name: sp_bacteria:Q46767

seq_documentation_block:

ID Q46767 PRELIMINARY; PRT; 732 AA.

AC Q46767;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

```

OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=49;
RX MEDLINE=96038471; PubMed=7480171;
RA Lim D.;
RT "Analysis of a retron EC86 and EC67 insertion site in Escherichia
RL plasmid 34:58-61(1995).
DR EMBL: U23723: AAA64865.1;
SQ SEQUENCE 732 AA; 81092 MW; D4576038D605D6D8 CRC64;

alignment_scores:
  Quality: 106.50      Length: 466
  Ratio: 0.475        Gaps: 25
Percent Similarity: 48.069 Percent Identity: 22.747

alignment_block:
US-09-303-518d-131 x Q46767 ..

Align seg 1/1 to: Q46767 from: 1 to: 732

28  CTGCCCATCGCGGAGACCGGAGCAAGTCATTATGACGGCCGCGCAT 77
|||||: : : : |||||
57  LeuProLeuAlaMetProGlu.....GlyLysAlaVal 68
78  TACCGAAGTC.....CGCTGCTGGCGAAGAATAT..... 108
|||||: : : : |||||: : : : |||||
68  lAlaLysPheValLeuHisProAlaPheLeuAlaLysSerTyPheProv 85
|||||: : : : |||||
109 .....GTGGCATGCGCCCTCG 126
85  aIserLeuPheGluArgPheSerLeuGluSerIleGlySerLys...Ala 100
127  ATGAAATCAAGGAGGCGGAGCGGCAAGGCAAGGCAAGTCTGTGA 176
|||||: : : : |||||
101  ValLysValLysProArgLysAspLysLys.....Ar 112
177  AGACAAAAGAAATCCGGCGGTAGTATTACTGCGCGGCTTCAGGCAAAA 226
|||||: : : : |||||
112  gGlyArgLysGluLysThrAlaCysLysIleValSerGlyLysL 129
227  TCCTGCTATTACCGTGGCGGAAACGGCTACTTCAGTCAGTCGATG 276
|||||: : : : |||||
129  yGluSerPheProSerPheLeuAspSerValAsnGlnAspThrLeu 145
277  GCGGTTGAAGGCAACGACGAAATCGAGTTCGACGCTAC..... 315
146  LysGlyGlnGlnAsnAspPheValThrLeuGluAsnIleSerIleLeu 162
316  .GTACTCGGAGCGCTGGCAAAATTTAGACGCGGAAAGTGGCGGCAACC 364
|||||: : : : |||||
162  uValSerAspLysValLysThrIleAsnSerAsnGluVal..... 175
365  TGATTAATCAGGCTTATGAGTCGCTTCGACCGCTCGCTTCAGCAAA 414
176 .....MetSerIleGluValAlaLeuHisThr...ProAspThrSer 188
415  ATCCCTGCGCTAGATGCC...GAGCGCTTGCGC..... 444
189  SerSerIleValAspSerPheGluValPheAlaSerGlnAsnGlyAla 205
445 .....ATCTTCGTCATGCGATGGACCAATCCGC 475
205  lIleAspLysAlaArgSerIleLysValLysGlyLeuThrPheMetPro 222
476  TGGCTGCGGACCTTACGCTCATCAAGAGCGCGGCAAGACTTCAAA 525
222  lLysAlaSerLysAspValAlaLeuLysValAlaGluPheSerPheLeu 238
526  CGCGGCTGTGTGATTGAGC.....CGCTGACCGAAGCTAAATC... 567

```

```

||| ||| ||| |||||: |||
239 ArgThrLeuArgGluLeuProGluLeuArgLeuSerGluProValIleSe 255
568 ....CATGTGTAAAGCAGCAGCGGCGACAGCTGCGCTCTGAAATGCTG 613
||| : : : : |||||: |||
255 rArgSerValIleGlnThrSerAsnLeuSerLeuProSerGluGlyAlaVal 272
614 CCAAT.....ATCGAAACACATGAATTTGGCGGCGCCGATCCTGCGCGC 657
: ||| : : : : ||| |||
272 alAsnProHisIleLysValAlaIlePheAspGly..... 283
658 TTGAGTGGCAGCAGCATTTCATTCAGCGCAGTCGCGCGCAATAAAC 707
||| : : : : |||
284 .....GlyLeuGlyIleAspPheAsnPro..... 292
708 CGTGTGGACCATCAATATCAACAGCTGATTGCTATCGGACGTTTGTTCG 757
|||: : : |||
293 ....TTPValThrGluThr.....P 299
758 TACAGGCGCTCTGAATACGAGCGCTGCTGCTTGGCGGCGCTGCAA 807
: |||: : : |||
299 heAsnGlyAsnAlaSerThrAsnAlaLysLeuSerHisGlyGlnAsp 315
808 GTCACAAACCGCGCTCTTGGTACCGCTTTGGTCCGAGGTGCTCA 857
|||||: : : : |||||: : : : |||||
316 ValThrSer...ThrValLeuPheGlyValLeuGlySerGluThrGlu 331
858 ACTTACCGCGCGGCAATTTGACGCGGACAAACCGCGCTGATTTCCGGTT 907
|||||: : : : |||||: : : : |||||
331 sLeuSerValProTyrcysAsnIleAspHisTyArgValLeuAspSerA 348
908 CGGTATTGAACGCTGCGATTCACAGCGCGCATGATTATTGCGACGC 957
|||||: : : : |||||
348 snValAsnAsnSerAspVal.....AspLeuPheAspValLeuIleArg 362
958 TACCACATCAGATTTCGCTTATCGAAGAGCGCGCAGCAAGAGCTGTT 1007
|||||: : : : |||||: : : : |||||
363 lIleLys.....SerValLeuGluGlnLysLysTyrrAspTyrrIleAs 376
1008 CGGCTGGTTCGCGCGCAG.....CCGACAAATACCTCCATCA 1045
: |||: : : |||
376 nLeuSerLeuGlyProArgLeuProValAspAspAspValHisValT 393
1046 CGCGCAGCTCTCGGCTTCTTCTTAAACAACTCTTCAAGTTCACG 1095
393 rPrThrSerLeuGluGluIleLeuAlaSerGlyGluThrLeuCysThr 409
1096 ACAGCGCTCAACGCGCGCGCATGCTACCGCATCGGCACTTATGA 1145
: |||: : : |||
410 ValAlaVal.....GlyAsnAspGlyGlnLeuPro...AlaLysLeuAs 423
1146 GCGGTAATGCGGTTGGACATCTCTGCTTCTTGGCGGATTTAA 1195
|||||: : : : |||||
423 nArgIleGlnPro.....ProAlaAspLeuValAsnGlyLeuS 436
1196 TCGTGGGATACCGACAGCGCGGCTTGGTGTGCTTGGATTTGAC 1245
|||||: : : |||
436 exValGlyAlaAlaThrSerLeu..... 443
1246 GAAGAGACCTCGCTTGTGCGAGTTCGCTGTC.....CCGGGCAA 1287
: |||: : : |||
444 SerAspSerTrpGluArgCysSerTyrrSerCysIleGlyProGlyArg 459

seq_name: sp_bacteriap:051099
seq_documentation_block:
ID_051099 PRELIMINARY; PRT; 773 AA.
AC_051099;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE HYPOTHETICAL 89.8 KDA PROTEIN.
GN BB0072.

```


163 LeuMetAlaGluLysLeuGluGluIlePheSerGlyPheGlnIleAsn

AW DNA-BINDING; nuclear protein; receptor; Transcription


```

1130 .....GTTACCATGGCGCGTTCGCGCGTGG 1104
306 roSerPheThrPheThrLeuProAlaAlaProAlaSerProThr 322
1103 ACGGCTGTCGTGAAC.....TTGAAGATTGTTTGTAG 1069
323 SerLeuLeuAlaProSerThrAsnProLeuLeuGluSerLeuLysMe 339
1068 GAATGGCGGAGAGTGGTGGCGTATGAGTATTGTCGGCGCGCGG 1019
339 tGlnThrProSerLeuProProCysProGluSerAlaAlaLat 356
1018 CAACCCAGCCGACGAC.....TCTTTGCTGCGGCTTCT 984
356 hrThrGluAlaLeuSerProProLysThrProSerLeuProProLeu 372
983 TCGATAACGGAAATCTGATTGTTGTTAGTCCCAATAATCATCGCGGC 934
373 GlyLeuSerGln.....SerGlyPr 379
933 TTGTGCAATCCGACCG.....TTCAATACCGAACCGGAATCACGC 893
379 oProGlyLeuLeuProSerProSerPheAspSerLysProProThrL 396
892 GGTGTGTC.....CGCTCAACCAATTCCGCGCGCGGTGAAGTTGAGAC 852
396 euLeuGlyLeuLeuProAlaProSerMetValProAlaThr.....Asp 410
851 ACCTTCGCACCCAAACGGTACGCAAGCGCGGTTGTTGACTTCAG 802
411 ThrLysAlaProProThrLeuGlnAlaGlu.....ThrAlaLath 423
801 GCCGCCCAAGCAACACCG.....CGCTCGGTATTC 770
423 rLysProGlnAlaThrSerAlaProSerProAlaProLysGlnSerPheL 440
769 GACGGCTGTTCAGCAACACCGTCCGATA.....GCAATC 735
440 euPheGlyThrGlnAsnThrSerProSerProAlaAlaProAlaAla 456
734 ACGTCTTGATAATTGATGTTCCACAGGTTTATTGCGCGCGGCTGCTC 685
457 SerSerAlaProProMetPheLysProIlePheThrAlaProProLysSe 473
684 GATGAATGAATGTCGTGCGTCAACGCG..... 654
473 rGluLysGluGlyProThrProProGlyProSerValThrAlaThrAlaP 490
654 ..... 654
490 roSerSerSerLeuProThrThrThrSerThrThrAlaProThrPhe 506
653 .....GCAGATGCGGCGCCCAATTCATGTTTCGATATT 616
507 GlnProValPheSerSerMetGlyProProAlaSer...ValProLeuPr 522
615 GCAGCATTTTCAGACGGCAGCTCTCGGCTGCTTTACACATGGA 566
522 oAlaProPheLysGlnThrThrThrProAlaThrAlaProThr..... 537
565 TTTTACGTTTCGGTTCAGCGCGCTCAATACCAACAGCGCGGTTGAAG... 519
538 .....ThrThrAlaProLeuPheThrGly 545
518 ...TCTTCGGCGGCTTCTTTGATGATGACCGTAGGTGCGGACCGCGG 472
546 LeuAlaSerAlaThrSerAlaValAlaProIleThrSerAlaSerProSe 562
471 ATTGGTGTCTCATCGGATTGACCAAGATGCGGACCGC.....TCGGCAT 428
562 rThrAspSerAlaSerLysProAlaPheGlyPheGlyIleAsnSerValS 579
427 CTACGGCAGGGATTTTGTGTAACGACGCGGTGCGAAGCGCAGTCCATAAG 378

```

```

1104 ..... 579 erSerSerSerValSerThrThrThrSerThrAlaThrAlaAlaSerGln 595
372 ..... 377 CCTGATTGAATCAGGTTCGCGCCACTTTTTCGCTGCTCAATTTTCCAG 328
328 ..... 596 Pro.....PheLeuPheGlyAlaProGlnAlaSerAl 606
606 ..... 327 CGTTCAGTACGTACGTTCGAACCTCGATTTTCGTCGTTG.....CCTT 284
284 ..... 606 aAlaSerPheThrProAlaMetGlySerIlePheGlnPheGlyLysProp 623
623 ..... 283 CAACGGCAATCAGCTGACTGAGTACGCGCTTTTCGCCACGGTGAATA 234
234 ..... 623 roAlaLeuProThrThrThrValThrPheSerGlnSer...Leu 638
638 ..... 233 CGCGCGATTTCGCTGAAGCGCGCGCAGTAAATACATACG.....CCCGG 190
190 ..... 639 HisThrAlaValProThrAlaThrSerSerSerAlaAlaAspPheSerGl 655
655 ..... 189 ATTCCTTTTGTCTTCAACAGCAGCTTGGCTTTTTCGCGGCTTTCACCTT 140
140 ..... 655 yPheGlySerThrLeuAlaThrSerAlaProAlaThrSerSerGlnProT 672
672 ..... 139 CTTTGATTTTCATCGAGGGCGCATCGCACATATTCTTCGCCAAGCAAC 90
90 ..... 672 hrLeuThrPheSerAsnThrSerThrProThrPheAsnIleProPheGly 688
688 ..... 89 GCGACTTCGGTAATGCGCGCGCTGATAAATGACTTGTCTCGGCTCTGCC 40
40 ..... 689 SerSerAlaLysSerProLeuProSerTyrProGlyAlaAsnProGlnPr 705
705 ..... 39 CGCGATGGGC 30
30 ..... 705 oAlaPheGly 708
708 .....
seq_name: sp_invertebrate:Q95SX4
seq_documentation_block:
ID Q95SX4 PRELIMINARY; PRT; 632 AA.
AC Q95SX4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE LD47460P.
GN CG4453.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Y, CN BW SP;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Parfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY060443; AAL25482.1;
SQ SEQUENCE 632 AA; 61593 MW; 781D83A42C6557E3 CRC64;

```

alignment_scores:

Quality: 106.00 Length: 538
Ratio: 0.453 Gaps: 19
Percent Similarity: 43.494 Percent Identity: 20.260

alignment_block:

US-09-303-518D-131/rev x Q95SX4 ..

Align seg 1/1 to: Q95SX4 from: 1 to: 632

```

1343 CAGCCTTCTCTCCATGTTTCCAGCACTTTGCGCAACAGCGGCGGTA 1294
    : : : : : : : : : : : : : : : : : : : : : : : :
46  LysProSerValGluThrThrThrProAlaProThrSerAlaGlnPh 62
1293 TTCGTATTTCGGCGGCGAGAGCTGCACAAAGCGAGCTCTCTTCGT 1244
    : : : : : : : : : : : : : : : : : : : : : : : :
62  eSerPheGlyPheGlyGlnSerAsnGlnGlyLysAspValAlaAspSerL 79
1243 CCAATTCACCAACCAACGCTGCGGCTGCTGCGTATGCGCGCAGGATT 1194
    : : : : : : : : : : : : : : : : : : : : : : : :
79  yLysThrGluAlaProLysThrPheMetPheGlyValSer..... 92
1193 AAATCGCCAAACAGAGTGGCAGGATGTCACAGCGATTACGCGCTC 1144
    : : : : : : : : : : : : : : : : : : : : : : : :
93  LysValGluGluProLysThrValSerPheGlyThrGlyIleLysGluTh 109
1143 ATAGTCGCGATCGGTACCATGCGCGCTGCGCGGTGTCGCGTTCG 1094
    : : : : : : : : : : : : : : : : : : : : : : : :
109 rThrAlaThrSerSerThrGluAlaThrAlaProThrProAlaAlaAla 126
1093 TGAACCTGAAGAGTTGTTTTAGGAAATGCCGAGAGTGGTGGCGGTG 1044
    : : : : : : : : : : : : : : : : : : : : : : : :
126 laProAlaProValGlnPheValPheLysAlaProThrThrAlaThrThr 142
1043 ATGGAGTATTGTCC..... 1029
143 AlaSerSerLeuThrThrThrIleSerThrThrSerAsnAlaProAlaLe 159
1028 ....GGCTGCGCGCAACCCAGCGAGCTCTTGTGCGGCTTCTT 983
159 uGlyGlyPheSerPheGlyAlaProSerSerSerThrValSerSerS 176
982 CGATACGGAATCTGATGTGGTAGCTGCCAATAATCATGCGCGCT 933
176 exThrThrSerThr.....SerAlaAsnPro 184
932 TGTGAATCGCACCGTCT..... 915
185 AlaAlaValLysProMetPheSerThrPheSerGlyAlaGlySerAlaVal 201
914 .....ATACGGAACCG..... 903
201 rSerThrSerSerSerGlnProValAlaLysAlaProThrLeuGlyP 219
902 .....GAAATCAGCGGTGTCGCGCTCAACCAAT... 873
218 heGlyValSerSerSerThrValThrThrThrThrThrThrLysVal 234
872 .....TCGCGCGGTAAGTTGAGACACCTTCGCA..... 843
235 PheAlaPheThrProAlaSerGlyLeuAspProAlaAlaAlaThrSerAl 251
842 .....CCCAAA 836
251 aProAlaAlaGlyAlaGlyPheSerPheGlySerGlnSerLysProAla 268
835 CGGTACGC.....AAGAGCGCGGTGTTGTG 810
268 hrThrGlnAsnThrGlyThrPhePheGlyGlnProThrAlaValAla 284
809 ACTTGAGCGCGCCAGGCAACCGCGCTGGTATTACAGCGCGCTGT 760
285 ProAlaThrProThrAsnProSerValSerIlePheGlyAlaProAl 301
759 TAGCAACAACGTCGATACATACGCTTGTGATAATTGATGGTCCACA 710
301 aThrSerThrAlaSerThrSerValSerAlaThrThrThrSerThrS 318
709 CGGTT.....TTATTCGCGCGGCTGGCTCGATGAAATGA 675
318 hrAlaAsnAlaIleAlaSerSerPheAlaProThrSerThr..... 331
674 ATGTGCGTCCACTCAAGCGCGGAGGATGCGGCGCCCAAAATTCATGTGT 625
332 .....ProGlnLeuPheGlyAsnTrpGlyGluLysLysThrAspLe 345
624 TTCGATATTGGCAGCATTTTTCAGACGGCAGCTCGCGCTGCTGCTTAC 575
345 uthrThrPheGlyAlaSerSerGlySerGlyThrThrThrProSerP 362
574 ACACATGGATTTCGTTGCGGTGCGGCTCAATACCAACAGGCGCGCT 525
362 heGlyTrp.....SerSerAsnGlyAspAla 370
524 TTCAGTCTTCGCGGCTTCTTTTGATGATGACCGTAGGTCGGCAGCCAG 475
371 AlaLysSerAsnSerAlaAla.....ValGlySerAlaAla... 382
474 CGATTGGTGTCTTCATTCGCAATTACGAGATGCGGACGCTCGCATCTA 425
383 .....ValProSerSerSerAlaSerT 390
424 CGCGAGGATTTCGTGAACGAGCGGTGCGAAGCGCAGTCCATAAGCCT 375
390 hrMetAlaThrProIlePheGly..... 397
374 GATTGAATCAGTTCGCGCGCACTTTTTCGCTGCTCAATTTGCCAGCC 325
398 .....SerSerSerMetPheGlyProSerSerS 407
324 TTCAGTACGTACGTTTTCGAACCTCGATTTCGTGCTGCTTCCCT...TCAACGG 278
407 rSerAsnAsnThrThrSerThrThrThrSerLeuProPheGlySerA 424
277 CAATCAGCAGTACTGAAGTACGCGCTTTTTCGCCACGCGTGAATAGCGGG 228
424 laAlaThrThr.....AlaAla 429
227 ATTTGCTGACGCGCGGAGTAATACTACGCGCGGATTCTTTTGTGTC 178
430 ThrThrProAlaGlyGlyAsnAlaAlaLeuThrGlyLeuPheGlyAsnVa 446
177 TTCAACACAGC.....ACTTGGCCTTTT 155
446 lGlyAsnSerLeuAlaGlyValGlyAlaProValAlaThrThrProAla 463
154 TGACGCGCTTTCCTTCCTTTCATTCGCGGCGC..... 117
463 laThrAlaAlaProLeuThrAsnIlePheGlyAsnProThrProVal 479
116 .....ATGCGACATATTCTTCG.....CCAAGCAACGC 88
480 AlaAlaAlaProValPheGlySerGlySerThrIleProSerAlaGl 496
87 GACTTCGGTAATGCGCGCGCTCATTAATGACTTGTCTCGGCTGCGCG 38
496 yPheGlyAlaProAlaAlaAlaProLeuAlaProAlaLeuProG 513
37 CGATGGCGCAGATT 24
513 lYAlaPheAsnPh 517

```